

(translation of the front page of the priority document of
Japanese Patent Application No.2000-221128)

PATENT OFFICE
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the
following application as filed with this Office.

Date of Application: July 21, 2000

Application Number : Patent Application 2000-221128

Applicant(s) : Canon Kabushiki Kaisha

August 25, 2000

Commissioner,
Patent Office

Kouzo OIKAWA

Certification Number 2000-3067852



CFM 1969.05
p200-0183

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

2000年 7月21日

出 願 番 号

Application Number:

特願2000-221128

出 願 人

Applicant (s):

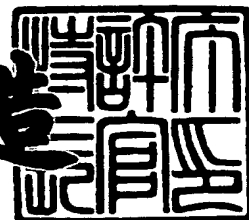
キヤノン株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年 8月25日

特 許 庁 長 官
Commissioner,
Patent Office

及 川 耕 造



出証番号 出証特2000-3067852

【書類名】 特許願

【整理番号】 4255023

【提出日】 平成12年 7月21日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00

【発明の名称】 音声情報処理方法、装置および記憶媒体

【請求項の数】 48

【発明者】

 【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社
 社内

 【氏名】 山田 雅章

【特許出願人】

 【識別番号】 000001007

 【氏名又は名称】 キヤノン株式会社

【代理人】

 【識別番号】 100076428

 【弁理士】

 【氏名又は名称】 大塚 康德

 【電話番号】 03-5276-3241

【選任した代理人】

 【識別番号】 100101306

 【弁理士】

 【氏名又は名称】 丸山 幸雄

 【電話番号】 03-5276-3241

【選任した代理人】

 【識別番号】 100115071

 【弁理士】

 【氏名又は名称】 大塚 康弘

 【電話番号】 03-5276-3241

【先の出願に基づく優先権主張】

【出願番号】 平成11年特許願第220496号

【出願日】 平成11年 8月 3日

【手数料の表示】

【予納台帳番号】 003458

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0001010

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声情報処理方法、装置および記憶媒体

【特許請求の範囲】

【請求項 1】 複数の音声素片を保持する素片辞書を生成する音声情報処理方法であって、

音声素片を符号化するための符号化方法を、複数の符号化方法の中から選択する選択工程と、

選択された符号化方法を用いて前記音声素片を符号化する符号化工程と、

符号化された前記音声素片を素片辞書に格納する格納工程と、

を有することを特徴とする音声情報処理方法。

【請求項 2】 前記複数の符号化方法の一つは、他の符号化方法とは量子化ステップ数が異なることを特徴とする請求項 1 に記載の音声情報処理方法。

【請求項 3】 前記複数の符号化方法の一つは、他の符号化方法とは量子化コードブックが異なることを特徴とする請求項 1 に記載の音声情報処理方法。

【請求項 4】 前記複数の符号化方法の一つは、他の符号化方法とは符号化方式が異なることを特徴とする請求項 1 に記載の音声情報処理方法。

【請求項 5】 前記複数の符号化方法の一つは、 μ -law方式を用いた符号化方法、スカラ量子化方式を用いた符号化方法、線形予測符号化方式を用いた符号化方法の何れかであることを特徴とする請求項 1 に記載の音声情報処理方法。

【請求項 6】 前記選択工程は、一部の音声素片を符号化しないように制御することを特徴とする請求項 1 に記載の音声情報処理方法。

【請求項 7】 複数の音声素片を保持する素片辞書を生成する音声情報処理装置であって、

音声素片を符号化するための符号化方法を、複数の符号化方法の中から選択する選択手段と、

選択された符号化方法を用いて前記音声素片を符号化する符号化手段と、

符号化された前記音声素片を素片辞書に格納する格納手段と、

を有することを特徴とする音声情報処理装置。

【請求項 8】 前記複数の符号化方法の一つは、他の符号化方法とは量子化

ステップ数が異なることを特徴とする請求項 7 に記載の音声情報処理装置。

【請求項 9】 前記複数の符号化方法の一つは、他の符号化方法とは量子化コードブックが異なることを特徴とする請求項 7 に記載の音声情報処理装置。

【請求項 10】 前記複数の符号化方法の一つは、他の符号化方法とは符号化方式が異なることを特徴とする請求項 7 に記載の音声情報処理装置。

【請求項 11】 前記複数の符号化方法の一つは、 μ -law方式を用いた符号化方法、スカラ量子化方式を用いた符号化方法、及び線形予測符号化方式を用いた符号化方法の何れかであることを特徴とする請求項 7 に記載の音声情報処理装置。

【請求項 12】 前記選択手段は、一部の音声素片を符号化しないように制御することを特徴とする請求項 7 に記載の音声情報処理装置。

【請求項 13】 複数の音声素片を保持する素片辞書を用いて音声を合成する音声情報処理方法であって、

素片辞書から読み出した音声素片を復号するための復号方法を、複数の復号方法の中から選択する選択工程と、

選択された復号方法を用いて前記音声素片を復号する復号工程と、

復号された前記音声素片に基づいて音声を合成する音声合成工程と、
を有することを特徴とする音声情報処理方法。

【請求項 14】 前記複数の復号方法の一つは、他の符号化方法とは量子化ステップ数が異なることを特徴とする請求項 13 に記載の音声情報処理方法。

【請求項 15】 前記複数の復号方法の一つは、他の符号化方法と量子化コードブックが異なることを特徴とする請求項 13 に記載の音声情報処理方法。

【請求項 16】 前記複数の復号方法の一つは、他の符号化方法とは符号化方式が異なることを特徴とする請求項 13 に記載の音声情報処理方法。

【請求項 17】 前記複数の復号方法の一つは、 μ -law方式に対応する復号方法、スカラ量子化方式に対応する復号方法、及び線形予測符号化方式のいずれかに対応する復号方法であることを特徴とする請求項 13 に記載の音声情報処理方法。

【請求項 18】 前記選択工程は、一部の音声素片を復号しないように制御

することを特徴とする請求項 1 3 に記載の音声情報処理方法。

【請求項 1 9】 複数の音声素片を保持する素片辞書を用いて音声を合成する音声情報処理装置であって、

素片辞書から読み出した音声素片を復号するための復号方法を、複数の復号方法の中から選択する選択手段と、

選択された復号方法を用いて前記音声素片を復号する復号手段と、

復号された前記音声素片に基づいて音声を合成する音声合成手段と、
を有することを特徴とする音声情報処理装置。

【請求項 2 0】 前記複数の復号方法の一つは、他の符号化方法とは量子化ステップ数が異なることを特徴とする請求項 1 9 に記載の音声情報処理装置。

【請求項 2 1】 前記複数の復号方法の一つは、他の符号化方法と量子化コードブックが異なることを特徴とする請求項 1 9 に記載の音声情報処理装置。

【請求項 2 2】 前記複数の復号方法の一つは、他の符号化方法とは符号化方式が異なることを特徴とする請求項 1 9 に記載の音声情報処理装置。

【請求項 2 3】 前記複数の復号方法の一つは、 μ -law方式に対応する復号方法、スカラ量子化方式に対応する復号方法、線形予測符号化方式のいずれかに対応する復号方法であることを特徴とする請求項 1 9 に記載の音声情報処理装置。

【請求項 2 4】 前記選択手段は、一部の音声素片を復号しないように制御することを特徴とする請求項 1 9 に記載の音声情報処理装置。

【請求項 2 5】 複数の音声素片を保持する素片辞書を生成する音声情報処理方法であって、

音声素片を符号化するための符号化方法を、前記音声素片の種類に応じて設定する設定工程と、

設定された符号化方法を用いて前記音声素片を符号化する符号化工程と、

符号化された前記音声素片を素片辞書に格納する格納工程と、

を有することを特徴とする音声情報処理方法。

【請求項 2 6】 前記設定工程は、前記音声素片の種類が破裂音である場合とそうでない場合とで、前記音声素片に設定する符号化方法を変更することを特

徴とする請求項 2 5 に記載の音声情報処理方法。

【請求項 2 7】 前記設定工程は、前記音声素片の種類が破裂音である場合、前記音声素片を符号化しないように設定することを特徴とする請求項 2 5 に記載の音声情報処理方法。

【請求項 2 8】 前記設定工程は、前記音声素片の種類が無声音である場合とそうでない場合とで、前記音声素片に設定する符号化方法を変更することを特徴とする請求項 2 5 に記載の音声情報処理方法。

【請求項 2 9】 前記設定工程は、前記音声素片の種類が鼻音である場合とそうでない場合とで、前記音声素片に設定する符号化方法を変更することを特徴とする請求項 2 5 に記載の音声情報処理方法。

【請求項 3 0】 複数の音声素片を保持する素片辞書を生成する音声情報処理装置であって、

音声素片を符号化するための符号化方法を、前記音声素片の種類に応じて設定する設定手段と、

設定された符号化方法を用いて前記音声素片を符号化する符号化手段と、

符号化された前記音声素片を素片辞書に格納する格納手段と、

を有することを特徴とする音声情報処理装置。

【請求項 3 1】 前記設定手段は、前記音声素片の種類が破裂音である場合とそうでない場合とで、前記音声素片に設定する符号化方法を変更することを特徴とする請求項 3 0 に記載の音声情報処理装置。

【請求項 3 2】 前記設定手段は、前記音声素片の種類が破裂音である場合、前記音声素片を符号化しないように設定することを特徴とする請求項 3 0 に記載の音声情報処理装置。

【請求項 3 3】 前記設定手段は、前記音声素片の種類が無声音である場合とそうでない場合とで、前記音声素片に設定する符号化方法を変更することを特徴とする請求項 3 0 に記載の音声情報処理装置。

【請求項 3 4】 前記設定手段は、前記音声素片の種類が鼻音である場合とそうでない場合とで、前記音声素片に設定する符号化方法を変更することを特徴とする請求項 3 0 に記載の音声情報処理装置。

【請求項 3 5】 複数の音声素片を保持する素片辞書を用いて音声を合成する音声情報処理方法であって、

素片辞書から読み出す音声素片を復号するための復号方法を、前記音声素片の種類に応じて設定する設定工程と、

設定された復号方法を用いて前記音声素片を復号する復号工程と、

復号された前記音声素片に基づいて音声を合成する音声合成工程と、
を有することを特徴とする音声情報処理方法。

【請求項 3 6】 前記設定工程は、前記音声素片の種類が破裂音である場合とそうでない場合とで、前記音声素片に設定する復号方法を変更することを特徴とする請求項 3 5 に記載の音声情報処理方法。

【請求項 3 7】 前記設定工程は、前記音声素片の種類が破裂音である場合、前記音声素片を復号しないように設定することを特徴とする請求項 3 5 に記載の音声情報処理方法。

【請求項 3 8】 前記設定工程は、前記音声素片の種類が無声音である場合とそうでない場合とで、前記音声素片に設定する復号方法を変更することを特徴とする請求項 3 5 に記載の音声情報処理方法。

【請求項 3 9】 前記設定工程は、前記音声素片の種類が鼻音である場合とそうでない場合とで、前記音声素片に設定する復号方法を変更することを特徴とする請求項 3 5 に記載の音声情報処理方法。

【請求項 4 0】 複数の音声素片を保持する素片辞書を用いて音声を合成する音声情報処理装置であって、

素片辞書から読み出した音声素片を復号するための復号方法を、前記音声素片の種類に応じて設定する設定手段と、

設定された復号方法を用いて前記音声素片を復号する復号手段と、

復号された前記音声素片に基づいて音声を合成する音声合成手段と、
を有することを特徴とする音声情報処理装置。

【請求項 4 1】 前記設定手段は、前記音声素片の種類が破裂音である場合とそうでない場合とで、前記音声素片に設定する復号方法を変更することを特徴とする請求項 4 0 に記載の音声情報処理装置。

【請求項 4 2】 前記設定手段は、前記音声素片の種類が破裂音である場合、前記音声素片を復号しないように設定することを特徴とする請求項 4 0 に記載の音声情報処理装置。

【請求項 4 3】 前記設定手段は、前記音声素片の種類が無声音である場合とそうでない場合とで、前記音声素片に設定する復号方法を変更することを特徴とする請求項 4 0 に記載の音声情報処理装置。

【請求項 4 4】 前記設定手段は、前記音声素片の種類が鼻音である場合とそうでない場合とで、前記音声素片に設定する復号方法を変更することを特徴とする請求項 4 0 に記載の音声情報処理装置。

【請求項 4 5】 請求項 1 ～ 6 の何れかに記載の音声情報処理方法をコンピュータに実現させるための制御プログラムを格納することを特徴とする記憶媒体。

【請求項 4 6】 請求項 1 3 ～ 1 8 の何れかに記載の音声情報処理方法をコンピュータに実現させるための制御プログラムを格納することを特徴とする記憶媒体。

【請求項 4 7】 請求項 2 5 ～ 2 9 の何れかに記載の音声情報処理方法をコンピュータに実現させるための制御プログラムを格納することを特徴とする記憶媒体。

【請求項 4 8】 請求項 3 5 ～ 3 9 の何れかに記載の音声情報処理方法をコンピュータに実現させるための制御プログラムを格納することを特徴とする記憶媒体。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、複数の音声素片を保持する素片辞書を生成する音声情報処理方法及び装置と、素片辞書を用いて音声合成する音声情報処理方法及び装置、および記憶媒体に関するものである。

【 0 0 0 2 】

【従来の技術】

計算機を用いて音声を合成する音声合成技術では、素片辞書を用いている。この素片辞書には、音素、CV/VC、VCV等の単位（合成単位）の音声素片が格納されている。そして音声合成時には、その素片辞書から適切な音声素片を選択し、選択された音声素片を変形、接続することによって、所望の合成音声を生成する。これを説明するフローチャートが図15のフローチャートである。

【0003】

まずステップS131で、仮名漢字混じり文などで表現された発声内容を入力する。次にステップS132で、その入力された発声内容を解析し、素片記号列{p0, p1, ...} および韻律決定のためのパラメータを得る。そして次にステップS133に進み、素片時間長、基本周波数、パワーといった韻律を決定する。次に素片辞書検索ステップS134において、ステップS132の入力解析で得られた素片記号列{p0, p1, ...}、及びステップS133の韻律決定で得られた韻律に対して適切な素片{w0, w1, ...}を素片辞書から検索する。そして次にステップS135に進み、ステップS134の素片辞書検索で得られた素片{w0, w1, ...}を、ステップS133で決定された韻律に合うように変形して接続する。こうしてステップS136に進み、ステップS135における素片変形、接続処理の結果を基に、合成音声として出力する。

【0004】

また、音声合成における有効な手法として、波形編集方式がある。これは、声帯振動に同期して、波形の重畳、ピッチ間隔の変更などを施す方式である。この方式には、少ない演算量で自然に近い合成音声を生成できるという利点がある。このような方式を用いた場合、素片辞書は、検索のためのインデックス、各素片に対応する波形データ（素片データともいう）およびその補助情報から構成される。この際、素片辞書に登録される素片データの全ては、 μ -lawまたはADPCM (Adaptive Differential Pulse Code Modulation) を用いて符号化される場合が多い。

【0005】

【発明が解決しようとする課題】

上記従来例では、以下の問題点があった。

【 0 0 0 6 】

まず、素片辞書に登録する全ての素片データを μ -lawやA-lawのような符号化方式を用いて符号化する場合には、固定の量子化テーブルを用いて各素片データを非一様に量子化するために、十分な圧縮効率を得ることができないという問題がある。これは、あらゆる種類の音声素片に対して最低限の品質を維持できるように量子化テーブルを設計する必要があるためである。

【 0 0 0 7 】

一方、素片辞書に登録する全ての素片データをADPCMのような符号化方式を用いて符号化する場合には、適応的アルゴリズムの演算量により、復号化時の演算量が増加してしまうという問題がある。これは、復号化に多くの演算を要するのでは、波形編集方式の利点（処理量が少ない）が損なわれるためである。

【 0 0 0 8 】

本発明は上記従来例に鑑みてなされたもので、素片辞書に登録する音声素片の品質を劣化させることなく、素片辞書に必要な記憶容量を非常に効率的に削減する技術を提供することを目的とする。

【 0 0 0 9 】

また、本発明は上記従来例に鑑みてなされたもので、自然で高品質な合成音声を生成する技術を提供することを目的とする。

【 0 0 1 0 】

【課題を解決するための手段】

上記目的を達成するために本発明の音声情報処理方法は以下のような工程を備える。即ち、

複数の音声素片を保持する素片辞書を生成する音声情報処理方法であって、音声素片を符号化するための符号化方法を、複数の符号化方法の中から選択する選択ステップと、選択された符号化方法を用いて前記音声素片を符号化する符号化ステップと、符号化された前記音声素片を素片辞書に格納する格納ステップとを有することを特徴とする。

【 0 0 1 1 】

また本発明の記憶媒体は、上記音声情報処理方法をコンピュータに実現させる

ための制御プログラムを格納することを特徴とする。

【 0 0 1 2 】

また、本発明の音声情報処理装置は、

複数の音声素片を保持する素片辞書を生成する音声情報処理装置であって、音声素片を符号化するための符号化方法を、複数の符号化方法の中から選択する選択手段と、選択された符号化方法を用いて前記音声素片を符号化する符号化手段と、符号化された前記音声素片を素片辞書に格納する格納手段とを有することを特徴とする。

【 0 0 1 3 】

また本発明の音声情報処理方法は、

複数の音声素片を保持する素片辞書を用いて音声を合成する音声情報処理方法であって、素片辞書から読み出した音声素片を復号するための復号方法を、複数の復号方法の中から選択する選択ステップと、選択された復号方法を用いて前記音声素片を復号する復号ステップと、復号された前記音声素片に基づいて音声を合成する音声合成ステップとを有することを特徴とする。

【 0 0 1 4 】

また、本発明の記憶媒体は、上記音声情報処理方法をコンピュータに実現させるための制御プログラムを格納することを特徴とする。

【 0 0 1 5 】

また、本発明の音声情報処理装置は、

複数の音声素片を保持する素片辞書を用いて音声を合成する音声情報処理装置であって、素片辞書から読み出した音声素片を復号するための復号方法を、複数の復号方法の中から選択する選択手段と、選択された復号方法を用いて前記音声素片を復号する復号手段と、復号された前記音声素片に基づいて音声を合成する音声合成手段とを有することを特徴とする。

【 0 0 1 6 】

また本発明の音声情報処理方法は、

複数の音声素片を保持する素片辞書を生成する音声情報処理方法であって、音声素片を符号化するための符号化方法を、前記音声素片の種類に応じて設定する

設定ステップと、設定された符号化方法を用いて前記音声素片を符号化する符号化ステップと、符号化された前記音声素片を素片辞書に格納する格納ステップとを有することを特徴とする。

【 0 0 1 7 】

また、本発明の記憶媒体は、上記音声情報処理方法をコンピュータに実現させるための制御プログラムを格納することを特徴とする。

【 0 0 1 8 】

また本発明の音声情報処理装置は、

複数の音声素片を保持する素片辞書を生成する音声情報処理装置であって、音声素片を符号化するための符号化方法を、前記音声素片の種類に応じて設定する設定手段と、設定された符号化方法を用いて前記音声素片を符号化する符号化手段と、符号化された前記音声素片を素片辞書に格納する格納手段とを有することを特徴とする。

【 0 0 1 9 】

また本発明の音声情報処理方法は、

複数の音声素片を保持する素片辞書を用いて音声合成する音声情報処理方法であって、素片辞書から読み出す音声素片を復号するための復号方法を、前記音声素片の種類に応じて設定する設定ステップと、設定された復号方法を用いて前記音声素片を復号する復号ステップと、復号された前記音声素片に基づいて音声を合成する音声合成ステップとを有することを特徴とする。

【 0 0 2 0 】

また、本発明の記憶媒体は、上記音声情報処理方法をコンピュータに実現させるための制御プログラムを格納することを特徴とする。

【 0 0 2 1 】

また本発明の音声情報処理装置は、

複数の音声素片を保持する素片辞書を用いて音声を合成する音声情報処理装置であって、素片辞書から読み出した音声素片を復号するための復号方法を、前記音声素片の種類に応じて設定する設定手段と、設定された復号方法を用いて前記音声素片を復号する復号手段と、復号された前記音声素片に基づいて音声を合成

する音声合成手段とを有することを特徴とする。

【 0 0 2 2 】

【発明の実施の形態】

以下、添付図面を参照して本発明の好適な実施の形態を詳細に説明する。本実施の形態では、（１）素片辞書を作成する方法（素片辞書作成アルゴリズム）、（２）この素片辞書を用いて音声を合成する方法（音声合成アルゴリズム）の各々について詳細に説明する。

【 0 0 2 3 】

図１は、本発明の実施の形態に係る音声情報処理装置の概略機能構成を示すブロック図である。この音声情報処理装置を用いて、各実施の形態における素片辞書作成アルゴリズムと音声合成アルゴリズムとを実現する。

【 0 0 2 4 】

図１において、１００は中央処理装置（ＣＰＵ）で、数値演算及び各種制御処理を実行し、バス１０５を介して接続されている後述する各部の動作を制御する。１０１はＲＡＭ及びＲＯＭ等を備えた記憶装置で、中央処理装置１００により実行される各種制御プログラムやデータ等を記憶するとともに、中央処理装置１００による制御処理で必要となる各種データを一時的に格納する。１０２はハードディスク装置等の外部記憶装置で、素片データ１１１および素片辞書１１２を有する。この素片データ１１１は、素片辞書１１２に登録する前の状態（即ち、非圧縮状態）の音声素片を保持する記録領域である。１０３は出力装置で、例えば、各種のプログラムの動作状況などを表示するモニタ、音声合成された音声を出力するスピーカ等を備える。１０４はキーボードやマウス等を備える入力装置である。ユーザは、この入力装置１０４を用いて、素片辞書１１２を作成するためのプログラムを制御したり、素片辞書１１２を用いて音声を合成するためのプログラムを制御したり、音声合成したいテキスト（複数の文字列を含む）の入力したりすることができる。

【 0 0 2 5 】

以上の構成に基づいて、以下の各実施の形態における素片辞書作成アルゴリズムと音声合成アルゴリズムとについて詳細に説明する。

【 0 0 2 6 】

〔実施の形態 1〕

図 1 の音声処理装置を用いて本発明の実施の形態 1 における素片辞書作成アルゴリズム及び音声合成アルゴリズムについて説明する。

【 0 0 2 7 】

本実施の形態 1 では、素片辞書 1 1 2 に登録する音声素片ごとに、量子化ステップ数の異なる複数の符号化方法（具体的には、7ビット μ -law方式と 8ビット μ -law方式）を選択する例について説明する。なお、素片辞書 1 1 2 に登録する音声素片は、音素、半音素、ダイフォン（CV, VC等）、VCV（またはCVC）、これらの組み合わせ等からなる。

【 0 0 2 8 】

（素片辞書の作成）

図 2 は、本発明の実施の形態 1 における素片辞書作成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1 0 1 からこのプログラムを読み出し、以下に示す手順を実行する。

【 0 0 2 9 】

まずステップ S 2 0 1 では、外部記憶装置 1 0 2 の素片データ 1 1 1 に記憶されている N 個の素片データ（各素片データは非圧縮状態にある）のそれぞれを指示するインデックス i を“0”に初期化する。なお、このインデックス i は記憶装置 1 0 1 に設けられている。

【 0 0 3 0 】

次にステップ S 2 0 2 に進み、そのインデックス i が指示する i 番目の素片データ W_i を読み出す。ここでは、その読み出されたデータ W_i を、

$$W_i = \{x_0, x_1, \dots, x_{T-1}\}$$

とする。ここで、 T は W_i の時間長（単位はサンプル）である。

【 0 0 3 1 】

次にステップ S 2 0 3 に進み、ステップ S 2 0 2 で読み出した素片データ W_i を、7ビット μ -law方式を用いて符号化する。こうして符号化された結果を、

$$C_i = \{c_0, c_1, \dots, c_{T-1}\}$$

とする。

【0032】

次にステップS204に進み、ステップS203の7ビット μ -law符号化によって生じた符号化歪み ρ を計算する。ここでは、この符号化歪み ρ の尺度として、例えば、平均2乗誤差を用いる場合について説明する。なお、平均2乗誤差は、以下のように表すことができる。

【0033】

$$\rho = (1/T) \cdot \sum (x_t - \mu(7)^{-1}(c_t))^2$$

…式(1)

ここで、 $\mu(7)^{-1}()$ は、7ビット μ -law復号化関数を示す。なお、ここで、「 \sum 」は $t=0$ から $t=T-1$ までの総和を示す。

【0034】

次にステップS205に進み、ステップS204で求めた符号化歪み ρ が、予め定められた閾値 ρ_0 より大きいかどうかを判定する。もし、 $\rho > \rho_0$ であれば、7ビット μ -law方式による符号化では音素データ W_i の波形が歪んでしまうと判断してステップS206に進み、符号化方法を量子化ビット数の異なる8ビット μ -law方式に切り換える。一方、それ以外の場合には、ステップS207に進む。ステップS206では、ステップS202で読み出した素片データ W_i を、8ビット μ -law方式を用いて符号化する。こうして符号化された結果を、

$$C_i = \{c_0, c_1, \dots, c_{T-1}\}$$

とする。

【0035】

ステップS207では、素片データ W_i の符号化情報等を素片辞書112に書き込む。符号化情報の他には、素片データ W_i の復号に必要な情報を書き込む。この符号化情報は、素片データ W_i を符号化した符号化方法を特定する情報であり、

7ビット μ -law方式であれば、符号化情報は“0”

8ビット μ -law方式であれば、符号化情報は“1”

とする。

【 0 0 3 6 】

次にステップ S 2 0 8 に進み、何れか一方の符号化方式によって符号化された素片データ C_i を素片辞書 1 1 2 に書き込む。そしてステップ S 2 0 9 に進み、 N 個の素片データ全てに対して前述の処理を行ったか否かを判定する。ここで、 $i = N - 1$ が成立すれば本アルゴリズムを終了する。一方、そうでなければステップ S 2 0 でインデックス i に 1 を加え、再びステップ S 2 0 2 に進み、更新されたインデックス i が指定する素片データを読み出す。このような処理を N 個の素片データを全てに対して繰り返し実行する。

【 0 0 3 7 】

以上説明したように本実施の形態 1 の素片辞書作成アルゴリズムによれば、素片辞書 1 1 2 に登録する音声素片ごとに、7 ビット μ -law 方式で符号化するか、8 ビット μ -law 方式で符号化するかを選択することができる。このように構成することにより、素片辞書に登録する音声素片の品質を劣化させることなく、素片辞書に必要な記憶容量を非常に効率的に削減することができる。また、従来と同程度の記憶容量をもつ素片辞書に、従来よりも多くの種類の音声素片を登録することもできる。

【 0 0 3 8 】

なお、本実施の形態 1 では、上述の素片辞書作成アルゴリズムを記憶装置 1 0 1 に記憶されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【 0 0 3 9 】

(音声合成)

図 3 は、本発明の実施の形態 1 における音声合成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1 0 1 からこのプログラムを読み出し、以下に示す手順を実行する。

【 0 0 4 0 】

まずステップ S 3 0 1 では、入力装置 1 0 4 の有するキーボードやマウスを用いて、日本語、英語或はその他の言語からなる文字列を入力する。日本語の場合は、仮名漢字混じり文などで表現された文字列を入力する。次にステップ S 3 0 2 では、入力された文字列を解析し、この文字列の音韻系列とこの文字列の韻律を決定するためのパラメータとを求める。次にステップ S 3 0 3 では、ステップ S 3 0 2 で求めた韻律パラメータに基づいて継続時間長（声の長さを制御する韻律）、基本周波数（声の高さを制御する韻律）、パワー（声の強さを制御する韻律）等の韻律を決定する。

【 0 0 4 1 】

ステップ S 3 0 4 では、ステップ S 3 0 2 で求めた音韻系列とステップ S 3 0 3 で決定した韻律とに基づいて最適な素片系列を求める。そして、この素片系列に含まれる音声素片の一つを選択し、その選択した音声素片に対応する素片データとこの素片データに対応する符号化情報とを検索する。そして、素片辞書 1 1 2 がハードディスク等の記憶媒体に格納されているならば、符号化情報、素片データの記憶領域へ順次シークする。また、素片辞書 1 1 2 が R A M 等の記憶媒体に格納されているならば、符号化情報、素片データの記憶領域へポインタ（アドレスレジスタ）を順次移動する。

【 0 0 4 2 】

ステップ S 3 0 5 では、ステップ S 3 0 4 で検索した符号化情報を素片辞書 1 1 2 から読み出す。この符号化情報は、ステップ S 3 0 4 で検索した素片データの符号化方法を示し、

符号化情報が“0”であれば、7ビット μ -law方式

符号化情報が“1”であれば、8ビット μ -law方式

であることを示す。

【 0 0 4 3 】

次にステップ S 3 0 6 では、ステップ S 3 0 5 で読み出した符号化情報を調べ、“0”であれば7ビット μ -law方式に対応する復号方法を選択し、ステップ S 3 0 7 に進む。一方、“1”であれば8ビット μ -law方式に対応する復号方法を選択し、ステップ S 3 0 9 に進む。

【 0 0 4 4 】

ステップ S 3 0 7 では、ステップ S 3 0 4 で検索した素片データ（7ビット μ -law方式で符号化されている）を素片辞書 1 1 2 から読み出す。次にステップ S 3 0 8 に進み、7ビット μ -law方式で符号化されている素片データを復号化する。

【 0 0 4 5 】

一方、ステップ S 3 0 9 では、ステップ S 3 0 4 で検索した素片データ（8ビット μ -law方式で符号化されている）を素片辞書 1 1 2 から読み出す。次にステップ S 3 1 0 に進み、8ビット μ -law方式で符号化されている素片データを復号化する。

【 0 0 4 6 】

ステップ S 3 1 1 では、ステップ S 3 0 4 で求めた素片系列に含まれる全ての音声素片に対応する素片データを復号化したか否かを判定する。全ての素片データを復号化した場合には、ステップ S 3 1 2 に進み、まだ復号化していない素片データがある場合には、次の素片データを復号するためにステップ S 3 0 4 に進む。

【 0 0 4 7 】

ステップ S 3 1 2 では、ステップ S 3 0 3 で決定した韻律に基づいて、復号化した音声素片のそれぞれを変形して接続する（即ち、波形編集する）。次にステップ S 3 1 3 に進み、ステップ S 3 1 2 で求めた合成音声を出力装置 1 0 3 の有するスピーカから出力する。

【 0 0 4 8 】

以上説明したように実施の形態 1 の音声合成アルゴリズムによれば、所望の音声素片を 7 ビット μ -law方式、8 ビット μ -law方式の何れかに対応する復号方法によって復号することができる。このように構成することにより、自然で高品質な合成音声を生成することができる。

【 0 0 4 9 】

なお、実施の形態 1 では、上述の音声合成アルゴリズムを記憶装置 1 0 1 に記憶されているプログラムに基づいて実現する場合について説明したが、この素片

辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【 0 0 5 0 】

〔実施の形態 1 の変形 1〕

実施の形態 1 では、符号化歪みが所定の閾値よりも大きくなる素片データについては 8 ビット μ -law 方式で符号化する例について説明したがこれに限るものではない。8 ビット μ -law 方式で符号化した後に符号化歪みを求め、この符号化歪みが所定の閾値よりも大きくなる素片データについては、符号化せずに素片辞書に登録してもよい。このように構成することにより、非定常性の音声素片（例えば、有声摩擦音や破裂音に分類される音声素片）の品質の劣化を防止することができる。また、このようにして生成した素片辞書を用いて合成音声を生成することにより、自然で高品質な合成音声を生成することもできる。

【 0 0 5 1 】

〔実施の形態 1 の変形 2〕

実施の形態 1 では、符号化歪みに応じて 7 ビット μ -law 方式で符号化するか、8 ビット μ -law 方式で符号化するかを選択する例について説明したがこれに限るものではない。音声素片の種類（有声摩擦音、破裂音、鼻音、その他の有声音、無声音等）に応じて、7 ビット μ -law 方式で符号化するか、8 ビット μ -law 方式で符号化するか、符号化しないで素片辞書 1 1 2 に登録するかを選択するように構成してもよい。例えば、有声摩擦音および破裂音に分類される音声素片については符号化せずに素片辞書 1 1 2 に登録し、無声音および鼻音に分類される音声素片については 7 ビット μ -law 方式で符号化して素片辞書 1 1 2 に登録し、その他の有声音に分類される音声素片については 8 ビット μ -law 方式で符号化して素片辞書 1 1 2 に登録するように構成してもよい。

【 0 0 5 2 】

〔実施の形態 2〕

次に、図 1 の音声処理装置を用いて本発明の実施の形態 2 における素片辞書作成アルゴリズム及び音声合成アルゴリズムについて説明する。

【 0 0 5 3 】

実施の形態 2 では、素片辞書 1 1 2 に登録する音声素片ごとに、量子化コードブックの異なる複数の符号化方法を選択する例について説明する。なお、素片辞書 1 1 2 に登録する音声素片は、音素、半音素、ダイフォン（CV, VC 等），VCV（またはCVC），これらの組み合わせ等からなる。

【0054】

（素片辞書の作成）

図 4 は、本発明の実施の形態 2 における素片辞書作成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1 0 1 からこのプログラムを読み出し、以下に示す手順を実行する。

【0055】

まずステップ S 4 0 1 では、外部記憶装置 1 0 2 の素片データ 1 1 1 に記憶されている N 個の素片データ（各素片データは非圧縮状態にある）のそれぞれを指示するインデックス i を、“0”に初期化する。なお、このインデックス i は記憶装置 1 0 1 に設定されている。

【0056】

次にステップ S 4 0 2 に進み、インデックス i が指示する i 番目の素片データ W_i を読み出す。こうして読み出されたデータを、

$$W_i = \{x_0, x_1, \dots, x_{T-1}\}$$

とする。ここで、 T は W_i の時間長（単位はサンプル）である。

【0057】

次にステップ S 4 0 3 に進み、ステップ S 4 0 2 で読み出した素片データ W_i のスカラ量子化コードブック Q_i を作成する。具体的には、スカラ量子化コードブック Q_i を用いて符号化された素片データ W_i を復号化し、復号化されたデータ系列 $Y_i = \{y_0, y_1, \dots, y_{T-1}\}$ の平均 2 乗誤差が最小となるように（つまり、符号化歪み ρ が最小となるように）設計する。この場合、例えば、LBG 法などのアルゴリズムを用いることが可能である。このように構成することにより、符号化によって生じる音声素片の波形の歪みを最小限にすることができる。なお、平均 2 乗誤差は、以下のように表すことができる。

【0058】

$$\rho = (1/T) \cdot \sum (x_t - y_t)^2 \quad \dots \text{式 (2)}$$

ここで、「 \sum 」は $t=0$ から $t=T-1$ までの総和を示す。

【0059】

次にステップS404に進み、ステップS403で作成したスカラ量子化コードブック Q_i 等を素片辞書112に書き込む。量子化コードブック Q_i の他には、素片データ W_i の復号に必要な情報を書き込む。そしてステップS405に進み、ステップS403で作成した量子化コードブック Q_i を用いて素片データ W_i を符号化（スカラ量子化）する。

【0060】

ここでコードブック Q_i を

$$Q_i = \{q_0, q_1, \dots, q_{N-1}\} \quad (N \text{ は量子化ステップ})$$

とすると、 $x_t (\in W_i)$ に対応する符号 c_t は、以下のように表すことができる。

【0061】

$$c_t = \arg \min (x_t - q_n)^2 \quad (0 \leq n < N) \quad \dots \text{式 (3)}$$

次にステップS406に進み、ステップS405で符号化された素片データ $C_i (= \{c_0, c_1, \dots, c_{T-1}\})$ を素片辞書112に書き込む。そしてステップS407に進み、 N 個の素片データの全てに対して前述の処理を行ったか否かを判定する。ここで、 $i=N-1$ が成立すれば本アルゴリズムを終了する。一方、そうでなければステップS408でインデックス i に1を加え、再びステップS402に進み、更新されたインデックス i が指定する素片データを読み出す。このような処理を N 個の素片データの全てに対して繰り返し実行する。

【0062】

以上説明したように実施の形態2の素片辞書作成アルゴリズムによれば、素片辞書112に登録する音声素片ごとに量子化コードブックを作成し、作成した量子化コードブックを用いてその音声素片をスカラ量子化することができる。このように構成することにより、素片辞書に登録する音声素片の品質を劣化させるこ

となく、素片辞書に必要な記憶容量を非常に効率的に削減することができる。また、従来と同程度の記憶容量をもつ素片辞書に、従来よりも多くの種類の音声素片を登録することもできる。

【 0 0 6 3 】

なお、実施の形態 2 では、上述の素片辞書作成アルゴリズムを記憶装置 1 0 1 に記憶されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【 0 0 6 4 】

（音声合成）

図 5 は、本発明の実施の形態 2 における音声合成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1 0 1 からこのプログラムを読み出し、以下に示す手順を実行する。

【 0 0 6 5 】

まずステップ S 5 0 1 では、入力装置 1 0 4 の有するキーボードやマウスを用いて、日本語、英語あるいはその他の言語からなる文字列を入力する。日本語の場合は、仮名漢字混じり文などで表現された文字列を入力する。次にステップ S 5 0 2 では、入力された文字列を解析し、この文字列の音韻系列とこの文字列の韻律を決定するためのパラメータとを求める。次にステップ S 5 0 3 では、ステップ S 5 0 2 で求めた韻律パラメータに基づいて継続時間長（声の長さを制御する韻律）、基本周波数（声の高さを制御する韻律）、パワー（声の強さを制御する韻律）等の韻律を決定する。

【 0 0 6 6 】

ステップ S 5 0 4 では、ステップ S 5 0 2 で求めた音韻系列とステップ S 5 0 3 で決定した韻律とに基づいて最適な素片系列を求める。そして、この素片系列に含まれる音声素片の一つを選択し、選択された音声素片に対応するスカラ量子化コードブック、素片データを検索する。そして、素片辞書 1 1 2 がハードディスク等の記憶媒体に格納されているならば、スカラ量子化コードブック、素片デ

ータの記憶領域へ順次シークする。また、素片辞書 1 1 2 が R A M 等の記憶媒体に格納されているならば、スカラ量子化コードブック、素片データの記憶領域へポインタ（アドレスレジスタ）を順次移動する。

【 0 0 6 7 】

ステップ S 5 0 5 では、ステップ S 5 0 4 で検索したスカラ量子化コードブックを素片辞書 1 1 2 から読み出す。ステップ S 5 0 6 では、ステップ S 5 0 4 で検索した素片データを素片辞書 1 1 2 から読み出す。ステップ S 5 0 7 では、ステップ S 5 0 5 で読み出したスカラ量子化コードブックを用いて、ステップ S 5 0 6 で読み出した素片データを復号化する。

【 0 0 6 8 】

ステップ S 5 0 8 では、ステップ S 5 0 4 で求めた素片系列に含まれる全ての音声素片に対応する素片データを復号化したか否かを判定する。全ての素片データを復号化した場合にはステップ S 5 0 9 に進み、まだ復号化していない素片データがある場合には、次の素片データを復号するためにステップ S 5 0 4 に進む。

【 0 0 6 9 】

ステップ S 5 0 9 では、ステップ S 5 0 3 で決定した韻律に基づいて、復号化した音声素片のそれぞれを変形して接続する（即ち、波形編集する）。次にステップ S 5 1 0 に進み、ステップ S 5 0 9 で求めた合成音声を出力装置 1 0 3 が有するスピーカから出力する。

【 0 0 7 0 】

以上説明したように実施の形態 2 の音声合成アルゴリズムによれば、所望の音声素片をその音声素片に最適な量子化コードブックを用いて復号することができるので、自然で高品質な合成音声を生成することができる。

【 0 0 7 1 】

なお、実施の形態 2 では、上述の音声合成アルゴリズムを記憶装置 1 0 1 に記憶されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【 0 0 7 2 】

〔実施の形態 2 の変形 1〕

実施の形態 2 では、前述の実施の形態 1 と同様に、1 サンプル当りのビット数（即ち、スカラ量子化の量子化ステップ数）を素片データ毎に変更することも可能である。これは、実施の形態 2 の処理手順に次のように変更することによって実現することができる。まず、素片辞書作成アルゴリズムでは、図 4 におけるステップ S 4 0 4 の処理（スカラ量子化コードブックの書き込み）に先立って、量子化ステップ数を決定し、決定した量子化ステップ数をコードブックとともに素片辞書 1 1 2 に記録する。そして、音声合成アルゴリズムでは、ステップ S 5 0 5 の処理（スカラ量子化コードブックの読み出し）に先立って、量子化ステップ数を素片辞書 1 1 2 から読み出す。この量子化ステップ数は、前述の実施の形態 1 と同様に、符号化歪みを基準に決定することができる。

【 0 0 7 3 】

〔実施の形態 2 の変形 2〕

実施の形態 2 の音声合成アルゴリズムでは、ステップ S 5 0 5 において、素片データごとに作成されたスカラ量子化コードブックを選択する場合について説明したが、これに限るものではない。例えば、単語辞書 1 1 2 が予め保持する複数種類のスカラ量子化コードブックの中から、性能の最も良い（即ち、量子化歪みが最小となる）コードブックを選択するように構成してもよい。

【 0 0 7 4 】

〔実施の形態 2 の変形 3〕

実施の形態 2 では、符号化歪みが最小となるように量子化コードブックを設計し、設計した量子化コードブックを用いて素片データをスカラ量子化する例について説明したがこれに限るものではない。符号化歪みが所定の閾値よりも大きくなる素片データについては、符号化せずに素片辞書に登録してもよい。このように構成することにより、非定常性の音声素片（例えば、有声摩擦音や破裂音に分類される音声素片）の品質の劣化を防止することができる。また、このようにして生成した素片辞書を用いて合成音声を生成することにより、自然で高品質な合成音声を生成することもできる。

【 0 0 7 5 】

〔実施の形態 3〕

次に、図 1 の音声処理装置を用いて本発明の実施の形態 3 における素片辞書作成アルゴリズム及び音声合成アルゴリズムについて説明する。

【 0 0 7 6 】

前述の実施の形態 2 では、素片辞書 1 1 2 に登録する音声素片ごとに、量子化コードブックの異なる複数の符号化方法を選択する例について説明した。これに対して、実施の形態 3 では、複数の素片クラスタごとに、量子化コードブックの異なる複数の符号化方法を選択する例について説明する。なお、素片辞書 1 1 2 に登録する音声素片は、音素、半音素、ダイフォン（CV，VC等），VCV（またはCVC），これらの組み合わせ等からなる。

【 0 0 7 7 】

（素片辞書の作成）

図 6 は、本発明の実施の形態 3 における素片辞書作成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1 0 1 からこのプログラムを読み出し、以下に示す手順を実行する。

【 0 0 7 8 】

まずステップ S 6 0 1 で、外部記憶装置 1 0 2 の素片データ 1 1 1 に記憶されている N 個の素片データ（各素片データは非圧縮状態にある）の全てを読み出す。次にステップ S 6 0 2 に進み、これら全ての音声素片を複数（M 個とする）の素片クラスタにクラスタリングする。具体的には、各音声素片の波形形状の類似度に応じて M 個の素片クラスタを作成する。

【 0 0 7 9 】

次にステップ S 6 0 3 で、M 個の素片クラスタのそれぞれを指示するインデックス i を“0”に初期化する。次にステップ S 6 0 4 に進み、i 番目の素片クラスタ L_i に対するスカラ量子化コードブック Q_i を作成する。次にステップ S 6 0 5 に進み、ステップ S 6 0 4 で作成したコードブック Q_i を素片辞書 1 1 2 に書き込む。

【 0 0 8 0 】

次にステップ S 6 0 6 に進み、M 個の素片クラスタの全てに対して前述の処理を行ったか否かを判定する。ここで、 $i = M - 1$ が成立すれば（M 個の素片クラスタの全てについて処理が終了）ステップ S 6 0 8（図 7）に進む。一方、そうでなければステップ S 6 0 7 に進んで、クラスタインデックス i に 1 を加え、再びステップ S 6 0 4 に進み、次の素片クラスタのスカラ量子化コードブックを作成する。

【 0 0 8 1 】

M 個の素片クラスタの全てに対してスカラ量子化コードブックを作成した後、本アルゴリズムはステップ S 6 0 8 に進む。ステップ S 6 0 8 では、外部記憶装置 1 0 2 の素片データ 1 1 1 に記憶されている N 個の素片音声のそれぞれを指示するインデックス i を“0”に初期化する。次にステップ S 6 0 9 に進み、 i 番目の素片データ W_i に対するスカラ量子化コードブック Q_i を選択する。ここで選択されるスカラ量子化コードブック Q_i は、素片データ W_i が属する素片クラスタに対応する量子化コードブックである。

【 0 0 8 2 】

次にステップ S 6 1 0 に進み、ステップ S 6 0 9 で選択したスカラ量子化コードブックを指定する情報（コードブック情報）等を素片辞書 1 1 2 に書き込む。コードブック情報の他には、素片データ W_i の復号に必要な情報を書き込む。そしてステップ S 6 1 1 に進み、ステップ S 6 0 4 で作成したコードブック Q_i を用いて素片データ W_i を符号化する。そしてステップ S 6 1 2 に進み、ステップ S 6 1 1 で符号化された素片データ $C_i (= \{c_0, c_1, \dots, c_{T-1}\})$ を素片辞書 1 1 2 に書き込む。

【 0 0 8 3 】

次にステップ S 6 1 3 に進み、N 個の素片データ全てに対して前述の処理を行ったか否かを判定する。もし $i = N - 1$ が成立すれば本アルゴリズムを終了する。一方、そうでなければステップ S 6 1 4 でインデックス i に 1 を加え、再びステップ S 6 0 9 に進み、次の素片データのスカラ量子化コードブックを作成する。

【 0 0 8 4 】

以上説明したように実施の形態 3 の素片辞書作成アルゴリズムによれば、複数の素片クラスタごとに、量子化コードブックの異なる複数の符号化方法を選択することができるので、素片辞書 1 1 2 に登録する量子化コードブックの数を前述の実施の形態 2 に比べて少なくすることができる。このように構成することにより、素片辞書に登録する音声素片の品質を劣化させることなく、素片辞書に必要な記憶容量を非常に効率的に削減することができる。また、従来と同程度の記憶容量をもつ素片辞書に、従来よりも多くの種類の音声素片を登録することもできる。

【 0 0 8 5 】

なお、実施の形態 3 では、上述の素片辞書作成アルゴリズムを記憶装置 1 0 1 に記憶されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【 0 0 8 6 】

(音声合成)

図 8 は、本発明の実施の形態 3 における音声合成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1 0 1 からこのプログラムを読み出し、以下に示す手順を実行する。なお、実施の形態では説明を簡単にするために、全素片クラスタに対応するコードブックが予め記憶装置 1 0 1 に読込まれているものとする。

【 0 0 8 7 】

ステップ S 8 0 1 ～ S 8 0 3 のそれぞれは、図 5 の S 5 0 1 ～ S 5 0 3 のそれぞれと同様の機能及び処理をするものとし、その説明を省略する。

【 0 0 8 8 】

ステップ S 8 0 4 では、ステップ S 8 0 2 で求めた音韻系列とステップ S 8 0 3 で決定した韻律とに基づいて最適な素片系列を求める。そして、この素片系列に含まれる音声素片の一つを選択し、選択された音声素片に対応するコードブッ

ク情報と素片データとを検索する。そして、素片辞書 1 1 2 がハードディスク等の記憶媒体に格納されているならば、コードブック情報の記憶領域および素片データの記憶領域へ順次シークする。また、素片辞書 1 1 2 が R A M 等の記憶媒体に格納されているならば、コードブック情報の記憶領域および素片データの記憶領域へポインタ（アドレスレジスタ）を順次移動する。

【 0 0 8 9 】

ステップ S 8 0 5 では、ステップ S 8 0 4 で検索したコードブック情報を読み出し、この音素データの素片クラスタとこの音声クラスタに対応するスカラ量子化コードブックとを決定する。ステップ S 8 0 6 では、素片辞書 1 1 2 を検索し、ステップ S 8 0 5 で決定したスカラ量子化コードブックを得る。ステップ S 8 0 7 では、ステップ S 8 0 4 で検索した素片データを素片辞書 1 1 2 から読み出す。ステップ S 8 0 8 では、ステップ S 8 0 6 で得られたスカラ量子化コードブックを用いて、ステップ S 8 0 7 で読み出した素片データを復号化する。

【 0 0 9 0 】

ステップ S 8 0 9 では、ステップ S 8 0 4 で求めた素片系列に含まれる全ての音声素片に対応する素片データを復号化したか否かを判定する。全ての素片データを復号化した場合には、ステップ S 8 1 0 に進み、まだ復号化していない素片データがある場合には、次の素片データを復号するためにステップ S 8 0 4 に進む。

【 0 0 9 1 】

ステップ S 8 1 0 では、ステップ S 8 0 3 で決定した韻律に基づいて、復号化した音声素片のそれぞれを変形して接続する（即ち、波形編集する）。次にステップ S 8 1 1 に進み、ステップ S 8 1 0 で求めた合成音声を出力装置 1 0 3 の有するスピーカから出力する。

【 0 0 9 2 】

以上説明したように実施の形態 3 の音声合成アルゴリズムによれば、所望の音声素片を、その音声素片の属する素片クラスタに最適な量子化コードブックを用いて復号することができるので、自然で高品質な合成音声を生成することができる。

【 0 0 9 3 】

なお、実施の形態 3 では、上述の音声合成アルゴリズムを記憶装置 1 0 1 に記憶されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【 0 0 9 4 】

〔実施の形態 3 の変形 1〕

実施の形態 3 の素片辞書作成アルゴリズムでは、音声素片の波形形状の類似度に応じて素片クラスタを作成する手順について説明したが、本発明はこれに限るものではない。音声素片の種類（有声摩擦音、破裂音、鼻音、その他の有声音、無声音等）に応じて素片クラスタを生成し、各素片クラスタごとに量子化コードブックを作成することも可能である。

【 0 0 9 5 】

〔実施の形態 3 の変形 2〕

実施の形態 3 の音声合成アルゴリズムでは、ステップ S 8 0 5 において、素片クラスタごとに作成されたスカラ量子化コードブックを選択する場合について説明したが、これに限るものではない。例えば、単語辞書 1 1 2 が保持する複数種類のスカラ量子化コードブックの中から、性能の最も良い（即ち、量子化歪みが最小となる）コードブックを選択するように構成してもよい。

【 0 0 9 6 】

〔実施の形態 3 の変形 3〕

実施の形態 3 では、ゲイン（パワー）を考慮してスカラ量子化を行うことも可能である。すなわち、ステップ S 6 0 9 において、スカラ量子化コードブックの選択に先んじて素片データのゲイン g を求める。次に、ステップ S 6 1 0 において、求めたゲイン g をコードブック情報とともに素片辞書 1 1 2 に書き込む。そして、ステップ S 6 1 1 において、ゲイン g を考慮した量子化を行う。これは前述の式（3）の代りに、以下の示す式を用いることを意味する。

【 0 0 9 7 】

$$ct = \arg \min (x_t - g \cdot q_n)^2 \quad (0 \leq n < N)$$

一方、音声合成アルゴリズムのステップ S 8 0 8（コードブックの参照）では、コードブック参照によって得られた値 q にゲイン g を乗じ、これを復号値とする。

【 0 0 9 8 】

〔実施の形態 3 の変形 4〕

実施の形態 3 では、素片クラスタごとに最適な量子化コードブックを設計し、設計した量子化コードブックを用いて各素片クラスタに属する素片データをスカラ量子化する例について説明したが、これに限るものではない。符号化による歪みが大きくなると判定された素片クラスタについては、符号化せずに素片辞書に登録してもよい。このように構成することにより、非定常性の音声素片（例えば、有声摩擦音や破裂音に分類される音声素片）の品質の劣化を防止することができる。また、このようにして生成した素片辞書を用いて合成音声を生成することにより、自然で高品質な合成音声を生成することもできる。

【 0 0 9 9 】

〔実施の形態 4〕

次に、図 1 の音声処理装置を用いて本発明の実施の形態 4 における素片辞書作成アルゴリズム及び音声合成アルゴリズムについて説明する。

【 0 1 0 0 】

実施の形態 4 では、素片データごとに線形予測係数と予測残差とを求め、求めた予測残差を最適な量子化コードブックで符号化する例について説明する。なお、素片辞書 1 1 2 に登録する音声素片は、音素、半音素、ダイフォンの（C V, V C 等）、V C V（または C V C）、これらの組み合わせ等からなる。

【 0 1 0 1 】

（素片辞書の作成）

図 9 は、本発明の実施の形態 4 における素片辞書作成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1 0 1 からこのプログラムを読み出し、以下に示す手順を実行する。

【 0 1 0 2 】

まずステップ S 9 0 1 では、外部記憶装置 1 0 2 の素片データ 1 1 1 に記憶されている N 個の素片データ（各素片データは非圧縮状態にある）のそれぞれを指示するインデックス i を“0”に初期化する。次にステップ S 9 0 2 では、インデックス i が指定する i 番目の音声素片の素片データ（符号化前の音声素片） W_i を読み出す。この読み出されたデータ W_i を、

$$W_i = \{x_0, x_1, \dots, x_{T-1}\}$$

とする。ここで、 T は W_i の時間長（単位はサンプル）である。

【0 1 0 3】

次にステップ S 9 0 3 に進み、ステップ S 9 0 2 で読み出した素片データ W_i の線形予測係数および予測残差を計算する。ここで、線形予測次数を L 次とすると、この線形予測モデルは、線形予測係数 a_l および予測残差 d_t を用いて以下のよう表される。

【0 1 0 4】

$$x_t = \sum_{l=1}^L a_l x_{t-l} + d_t \quad \dots \text{式 (4)}$$

ここで、 \sum は $l = 1 \sim L$ までの総和を示す。

【0 1 0 5】

そこで、予測残差 d_t の二乗和、

$$\sum d_t^2 \quad \dots \text{式 (5)}$$

を最小化する線形予測係数 a_l を決定する。ここで、 \sum は $t = 1 \sim T-1$ までの総和を示す。

【0 1 0 6】

次にステップ S 9 0 4 に進み、ステップ S 9 0 3 で求めた線形予測係数 a_l を素片辞書 1 1 2 に書き込む。そしてステップ S 9 0 5 に進み、ステップ S 9 0 3 で求めた予測残差 d_t の量子化コードブック Q_i を作成する。具体的には、量子化コードブック Q_i を用いて符号化された予測残差 d_t を復号化し、復号化されたデータ系列 $E_i = \{e_1, e_{l+1}, \dots, e_{T-1}\}$ の平均 2 乗誤差が最小となるように（つまり、符号化歪み ρ が最小となるように）設計する。この場合、例えば、L B G 法などのアルゴリズムを用いることが可能である。このように構成することにより、符号化によって生じる音声素片の波形の歪みを最小限にすることができる。

。なお、平均 2 乗誤差は、以下のように表すことができる。

【0 1 0 7】

$$\rho = (1/T) \cdot \sum (d_t - e_t)^2 \quad \dots \text{式 (6)}$$

ここで、 \sum は $t = 1 \sim T - 1$ までの総和を示す。

【0 1 0 8】

次にステップ S 9 0 6 に進み、ステップ S 9 0 5 で作成した量子化コードブック Q_i 等を素片辞書 1 1 2 に書き込む。このコードブック Q_i の他には、素片データ W_i の復号に必要な情報を書き込む。そしてステップ S 9 0 7 に進み、ステップ S 9 0 3 で求めた線形予測係数 a_l とステップ S 9 0 5 で作成したコードブック Q_i とを用いて素片データ W_i を線形予測符号化する。ここで、コードブック Q_i を、

$$Q_i = \{q_0, q_1, \dots, q_{N-1}\} \quad (N \text{ は量子化ステップ})$$

とすると、 $x_t (\in W_i)$ に対応する符号 c_t は、以下のように表すことができる。

【0 1 0 9】

$$c_t = \arg \min_n (x_t - \sum a_l y_{t-l} - q_n)^2 \quad (0 \leq n < N)$$

…式 (7)

ここで、 y_t は x_t を本手法で符号化した後、復号した値である。

【0 1 1 0】

次にステップ S 9 0 8 に進み、ステップ S 9 0 7 で符号化された素片データ $C_i (= \{c_0, c_1, \dots, c_{T-1}\})$ を素片辞書 1 1 2 に書き込む。そしてステップ S 9 0 9 に進み、 N 個の素片データの全てに対して前述の処理を行ったか否かを判定する。ここで、 $i = N - 1$ が成立すれば本アルゴリズムを終了する。一方、そうでなければステップ S 9 1 0 でインデックス i に 1 を加え、再びステップ S 9 0 2 に進み、更新されたインデックス i が指定する素片データを読み出す。このような処理を N 個の素片データの全てに対して繰り返し実行する。

【0 1 1 1】

以上説明したように実施の形態 4 の素片辞書作成アルゴリズムによれば、素片辞書 1 1 2 に登録する音声素片ごとに線形予測係数と予測残差とを求め、求めた

予測残差を最適な量子化コードブックで符号化することができる。このように構成することにより、素片辞書に登録する音声素片の品質を劣化させることなく、素片辞書に必要な記憶容量を非常に効率的に削減することができる。また、従来と同程度の記憶容量をもつ素片辞書に、従来よりも多くの種類の音声素片に登録することもできる。

【0 1 1 2】

なお、実施の形態 4 では、上述の素片辞書作成アルゴリズムを記憶装置 1 0 1 に記憶されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【0 1 1 3】

(音声合成)

図 1 0 は、本発明の実施の形態 4 における音声合成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1 0 1 からこのプログラムを読み出し、以下に示す手順を実行する。

【0 1 1 4】

まずステップ S 1 0 0 1 では、入力装置 1 0 4 の有するキーボードやマウスを用いて、日本語、英語あるいはその他の言語からなる文字列を入力する。日本語の場合は、仮名漢字混じり文などで表現された文字列を入力する。次にステップ S 1 0 0 2 では、入力された文字列を解析し、この文字列の音韻系列とこの文字列の韻律を決定するためのパラメータとを求める。次にステップ S 1 0 0 3 では、ステップ S 1 0 0 2 で求めた韻律パラメータに基づいて継続時間長（声の長さを制御する韻律）、基本周波数（声の高さを制御する韻律）、パワー（声の強さを制御する韻律）等の韻律を決定する。

【0 1 1 5】

ステップ S 1 0 0 4 では、ステップ S 1 0 0 2 で求めた音韻系列とステップ S 1 0 0 3 で決定した韻律とに基づいて最適な素片系列を求める。そして、この素片系列に含まれる音声素片の一つを選択し、選択された音声素片に対応する線形

予測係数、量子化コードブック、予測残差を検索する。そして、素片辞書 1 1 2 がハードディスク等の記憶媒体に格納されているならば、線形予測係数、量子化コードブック、予測残差の記憶領域へ順次シークする。また、素片辞書 1 1 2 が R A M 等の記憶媒体に格納されているならば、線形予測係数、量子化コードブック、予測残差の記憶領域へポインタ（アドレスレジスタ）を順次移動する。

【 0 1 1 6 】

ステップ S 1 0 0 5 では、ステップ S 1 0 0 4 で検索した予測係数を素片辞書 1 1 2 から読み出す。ステップ S 1 0 0 6 では、ステップ S 1 0 0 4 で検索した量子化コードブックを素片辞書 1 1 2 から読み出す。ステップ S 1 0 0 7 では、ステップ S 1 0 0 4 で検索した予測残差を素片辞書 1 1 2 から読み出す。ステップ S 1 0 0 8 では、予測係数、量子化コードブック、直前のサンプルの復号化データを用いて予測残差を復号化し、素片データを求める。

【 0 1 1 7 】

ステップ S 1 0 0 9 では、ステップ S 1 0 0 4 で求めた素片系列に含まれる全ての音声素片に対応する素片データを復号化したか否かを判定する。全ての素片データを復号化した場合には、ステップ S 1 0 1 0 に進み、まだ復号化していない素片データがある場合には、次の素片データを復号するためにステップ S 1 0 0 4 に進む。

【 0 1 1 8 】

ステップ S 1 0 1 0 では、ステップ S 1 0 0 3 で決定した韻律に基づいて、復号化した音声素片のそれぞれを変形して接続する（即ち、波形編集する）。次にステップ S 1 0 1 1 に進み、ステップ S 1 0 1 0 で求めた合成音声を出力装置 1 0 3 の有するスピーカから出力する。

【 0 1 1 9 】

以上説明したように実施の形態 4 の音声合成アルゴリズムによれば、所望の音声素片をその音声素片に最適なコードブックを用いて復号することができるので、自然で高品質な合成音声を生成することができる。

【 0 1 2 0 】

なお、実施の形態 4 では、上述の音声合成アルゴリズムを記憶装置 1 0 1 に記

億されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【 0 1 2 1 】

〔実施の形態 4 の変形 1〕

実施の形態 4 では、前述の実施の形態 1 と同様に、1 サンプル当りのビット数（即ち、量子化ステップ数）を素片データ毎に変更することも可能である。これは、実施の形態 4 の処理手順に次のように変更することによって実現することができる。まず、素片辞書作成アルゴリズムでは、ステップ S 9 0 5 の処理（量子化コードブックの書き込み）に先立って、量子化ステップ数を決定し、決定した量子化ステップ数をコードブックとともに素片辞書 1 1 2 に記録する。そして、音声合成アルゴリズムでは、ステップ S 1 0 0 6 の処理（量子化コードブックの読み出し）に先立って、量子化ステップ数を素片辞書 1 1 2 から読み出す。この量子化ステップ数は、前述の実施の形態 1 と同様に、符号化歪みを基準に決定することができる。

【 0 1 2 2 】

〔実施の形態 4 の変形 2〕

実施の形態 4 では、素片データごとに線形予測次数 L を変更することも可能である。これは、実施の形態 4 の処理手順に次のように変更することによって実現することができる。まず、素片辞書作成アルゴリズムでは、ステップ S 9 0 4 の処理（予測係数の書き込み）に先立って、予測次数を設定し、設定した予測次数を予測係数とともに素片辞書 1 1 2 に記録する。そして、音声合成アルゴリズムでは、ステップ S 1 0 0 5 の処理（予測係数の読み出し）に先立って、予測次数を素片辞書 1 1 2 から読み出す。この予測次数は、前述の実施の形態 1 と同様に、符号化歪みを基準に設定することができる。

【 0 1 2 3 】

〔実施の形態 4 の変形 3〕

実施の形態 4 では、ステップ S 9 0 5 で作成した量子化コードブックの符号化性能を更に向上させることも可能である。これは、ステップ S 9 0 5 では、予測

残差 d_t に対してコードブックを最適化するのに対し、ステップ S 9 0 7 では、

$$x_t - \sum a_l y_{t-l} \quad (\neq d_t = x_t - \sum a_l x_{t-l}) \quad \cdots \text{式 (8)}$$

に対して量子化コードブックを参照するためである。この量子化コードブックを更新するアルゴリズムとしては、A b S (Analysis by Synthesis) 法等を用いることができる。ここで Σ は、 $l = 1 \sim L$ の総和を示している。

【 0 1 2 4 】

〔実施の形態 4 の変形 4〕

実施の形態 4 では、1 つの素片データに対して 1 つの量子化コードブックを設計する場合について説明したが、これに限るものではない。複数の素片データに対して 1 つの量子化コードブックを設計してもよい。例えば、実施の形態 3 のように、N 個の素片データを M 個の素片クラスタにクラスタリングし、素片クラスタごとに量子化コードブックを設計するように構成してもよい。

【 0 1 2 5 】

〔実施の形態 4 の変形 5〕

前述の実施の形態 4 において、素片データの開始から L サンプルのデータについては符号化を行わず、そのまま素片辞書 1 1 2 に書き込んでも良い。これにより、素片データの開始から L サンプルについては線形予測がうまく行われれないという現象を回避することが可能である。

【 0 1 2 6 】

〔実施の形態 4 の変形 6〕

実施の形態 4 では、ステップ S 9 0 7 において、 x_t に関して最適となる符号 c_t を求めていたが、 x_t 以降の m サンプルも考慮にいれて最適となる符号 c_t を求めることも可能である。これは、符号 c_t を仮決めして再帰的に探索する（木構造探索する）ことにより実現できる。

【 0 1 2 7 】

〔実施の形態 4 の変形 7〕

実施の形態 4 では、符号化歪みが最小となるように量子化コードブックを設計し、設計した量子化コードブックを用いて素片データを線形符号化する例について説明したがこれに限るものではない。符号化歪みが所定の閾値よりも大きくな

る素片データについては、符号化せずに素片辞書に登録してもよい。このように構成することにより、非定常性の音声素片（例えば、有声摩擦音や破裂音に分類される音声素片）の品質の劣化を防止することができる。また、このようにして生成した素片辞書を用いて合成音声を生成することにより、自然で高品質な合成音声を生成することもできる。

【 0 1 2 8 】

〔実施の形態 5〕

次に、図 1 の音声処理装置を用いて本発明の実施の形態 5 における素片辞書作成アルゴリズム及び音声合成アルゴリズムについて説明する。

【 0 1 2 9 】

実施の形態 5 では、前述の各実施の形態で用いた様々な種類の符号化方式を組み合わせ、素片辞書 1 1 2 に登録する素片データごとに、最適な符号化方法を選択する例について説明する。実施の形態 5 では更に、非定常性の音声素片（例えば、有声摩擦音や破裂音に分類される音声素片）を非圧縮とする場合についても説明する。なお、素片辞書 1 1 2 に登録する音声素片は、音素、半音素、ダイフオン（C V，V C 等），V C V（または C V C），これらの組み合わせ等からなる。

【 0 1 3 0 】

（素片辞書作成）

図 1 1 は、本発明の実施の形態 5 における素片辞書作成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1 0 1 からこのプログラムを読み出し、以下に示す手順を実行する。

【 0 1 3 1 】

まずステップ S 1 1 0 1 では、外部記憶装置 1 0 2 の素片データ 1 1 1 に記憶されている N 個の素片データ（各素片データは非圧縮状態にある）のそれぞれを指示するインデックス i を、“0”に初期化する。なお、このインデックス i は記憶装置 1 0 1 に設定されている。

【 0 1 3 2 】

次にステップ S 1 1 0 2 に進み、インデックス i が指示する i 番目の素片データ W_i を読み出す。こうして読込まれたデータを、

$$W_i = \{x_0, x_1, \dots, x_{T-1}\}$$

とする。ここで、 T は W_i の時間長（単位はサンプル）である。

【 0 1 3 3 】

次にステップ S 1 1 0 3 に進み、ステップ S 1 1 0 2 で読み出した素片データ W_i を、前述の実施の形態 4 で説明した符号化方式（即ち、線形予測符号化方式）を用いて符号化する。

【 0 1 3 4 】

次にステップ S 1 1 0 4 に進み、この符号化方式による符号化歪み ρ を計算する。次にステップ S 1 1 0 5 に進み、ステップ S 1 1 0 3 で求めた符号化歪み ρ が、予め定められた閾値 ρ_0 より大きいかな否かを判定する。もし、 $\rho > \rho_0$ であれば、ステップ S 1 1 0 8 に進み、素片データ W_i を別の符号化方式を用いて符号化する。一方、 $\rho > \rho_0$ でなければ、ステップ S 1 1 0 6 に進む。

【 0 1 3 5 】

ステップ S 1 1 0 6 では、素片データ W_i の符号化情報を素片辞書 1 1 2 に書き込む。この符号化情報には、素片データ W_i を符号化した符号化方法を特定する情報や素片データ W_i の復号に必要な情報（予測係数、量子化コードブック等）などが含まれている。次にステップ S 1 1 0 7 に進み、ステップ S 1 1 0 3 で符号化された素片データ W_i を素片辞書 1 1 2 に書き込み、ステップ S 1 1 2 0 に進む。

【 0 1 3 6 】

一方、ステップ S 1 1 0 8 では、ステップ S 1 1 0 2 で読み出した素片データ W_i を、前述の実施の形態 1 で説明した符号化方式（即ち、7 ビット μ -law 方式又は 8 ビット μ -law 方式）を用いて符号化する。

【 0 1 3 7 】

次にステップ S 1 1 0 9 に進み、この符号化方式による符号化歪み ρ を計算する。次にステップ S 1 1 1 0 に進み、ステップ S 1 1 0 9 で求めた符号化歪み ρ が、予め定められた閾値 ρ_1 より大きいかな否かを判定する。もし、 $\rho > \rho_1$ であれ

ば、ステップ S 1 1 1 3 に進み、素片データ W_i を更に別の符号化方式を用いて符号化する。一方、 $\rho > \rho_1$ でなければ、ステップ S 1 1 1 1 に進む。

【 0 1 3 8 】

ステップ S 1 1 1 1 では、素片データ W_i の符号化情報を素片辞書 1 1 2 に書き込む。この符号化情報には、素片データ W_i を符号化した符号化方法を特定する情報や素片データ W_i の復号に必要な情報などが含まれている。次にステップ S 1 1 1 2 に進み、ステップ S 1 1 0 8 で符号化された素片データ W_i を素片辞書 1 1 2 に書き込み、ステップ S 1 1 2 0 に進む。

【 0 1 3 9 】

一方、ステップ S 1 1 1 3 では、ステップ S 1 1 0 2 で読み出した素片データ W_i を、前述の実施の形態 2 または 3 で説明した符号化方式（即ち、スカラ量子化方式）を用いて符号化する。次にステップ S 1 1 1 4 に進み、この符号化方式による符号化歪み ρ を計算する。次にステップ S 1 1 1 5 に進み、ステップ S 1 1 1 4 で求めた符号化歪み ρ が、予め定められた閾値 ρ_2 より大きいかなんかを判定する。例えば、非定常性の音声素片（例えば、有声摩擦音や破裂音に分類される音声素片）の場合は、波形形状が大きく変動するため、 $\rho > \rho_2$ とはならない。もし、 $\rho > \rho_2$ であれば、ステップ S 1 1 1 8 に進む。一方、 $\rho > \rho_2$ でなければ、ステップ S 1 1 1 6 に進む。

【 0 1 4 0 】

ステップ S 1 1 1 6 では、素片データ W_i の符号化情報を素片辞書 1 1 2 に書き込む。この符号化情報には、素片データ W_i を符号化した符号化方法を特定する情報や素片データ W_i の復号に必要な情報（量子化コードブック等）などが含まれている。次にステップ S 1 1 1 7 に進み、ステップ S 1 1 1 3 で符号化された素片データ W_i を素片辞書 1 1 2 に書き込み、ステップ S 1 1 2 0 に進む。

【 0 1 4 1 】

一方、ステップ S 1 1 1 8 では、ステップ S 1 1 0 2 で読み出した素片データ W_i を非圧縮とし、この素片データ W_i の符号化情報を素片辞書 1 1 2 に書き込む。この符号化情報には、素片データ W_i が符号化されていないことを示す情報などが含まれている。次にステップ S 1 1 1 9 に進み、この素片データ W_i を素片

辞書 1 1 2 に書き込み、ステップ S 1 1 2 0 に進む。このように構成することにより、非定常性の音声素片の品質の劣化を防止することができる。

【 0 1 4 2 】

ステップ S 1 1 2 0 では、N 個の素片データの全てに対して前述の処理を行ったか否かを判定する。ここで、 $i = N - 1$ が成立すれば本アルゴリズムを終了する。一方、そうでなければステップ S 1 1 2 1 でインデックス i に 1 を加え、再びステップ S 1 1 0 2 に戻り、更新されたインデックス i が指定する素片データを読み出す。このような処理を N 個の素片データの全てに対して繰り返し実行する。

【 0 1 4 3 】

以上説明したように実施の形態 5 の素片辞書作成アルゴリズムによれば、素片辞書 1 1 2 に登録する音声素片ごとに、 μ -law 方式で符号化するか、スカラ量子化方式で符号化するか、或は線形予測符号化方式で符号化するかを選択することができる。

【 0 1 4 4 】

このように構成することにより、素片辞書に登録する音声素片の品質を劣化させることなく、素片辞書に必要な記憶容量を非常に効率的に削減することができる。また、従来と同程度の記憶容量をもつ素片辞書に、従来よりも多くの種類の音声素片を登録することもできる。

【 0 1 4 5 】

なお、実施の形態 5 では、上述の素片辞書作成アルゴリズムを記憶装置 1 0 1 に記憶されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【 0 1 4 6 】

(音声合成)

図 1 2 は、本発明の実施の形態 5 における音声合成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1

0 1 からこのプログラムを読み出し、以下に示す手順を実行する。

【 0 1 4 7 】

まずステップ S 1 2 0 1 では、入力装置 1 0 4 の有するキーボードやマウスを用いて、日本語、英語あるいはその他の言語からなる文字列を入力する。日本語の場合は、仮名漢字混じり文などで表現された文字列を入力する。次にステップ S 1 2 0 2 では、入力された文字列を解析し、この文字列の音韻系列とこの文字列の韻律を決定するためのパラメータとを求める。次にステップ S 1 2 0 3 では、ステップ S 1 2 0 2 で求めた韻律パラメータに基づいて、継続時間長（声の長さを制御する韻律）、基本周波数（声の高さを制御する韻律）、パワー（声の強さを制御する韻律）等の韻律を決定する。

【 0 1 4 8 】

ステップ S 1 2 0 4 では、ステップ S 1 2 0 2 で求めた音韻系列とステップ S 1 2 0 3 で決定した韻律とに基づいて最適な素片系列を求める。そして、この素片系列に含まれる音声素片の一つを選択し、選択された音声素片に対応する素片データ、符号化情報を検索する。そして、素片辞書 1 1 2 がハードディスク等の記憶媒体に格納されているならば、素片データ、符号化情報の記憶領域へ順次シークする。また、素片辞書 1 1 2 が R A M 等の記憶媒体に格納されているならば、素片データ、符号化情報の記憶領域へポインタ（アドレスレジスタ）を順次移動する。

【 0 1 4 9 】

ステップ S 1 2 0 5 では、ステップ S 1 2 0 4 で検索した符号化情報を素片辞書 1 1 2 から読み出す。ステップ S 1 2 0 6 では、ステップ S 1 2 0 4 で検索した素片データを素片辞書 1 1 2 から読み出す。

【 0 1 5 0 】

次にステップ S 1 2 0 7 に進み、ステップ S 1 2 0 5 で読み出した符号化情報に基づいて、ステップ S 1 2 0 6 で読み出した素片データが符号化されているか否かを判別する。符号化されている場合には、符号化方法を特定するためにステップ S 1 2 0 8 に進む。符号化されていない場合には、ステップ S 1 2 1 5 に進む。

【 0 1 5 1 】

ステップ S 1 2 0 8 では、ステップ S 1 2 0 5 で読み出した符号化情報に基づいて、ステップ S 1 2 0 6 で読み出した素片データの符号化方法を調べる。線形予測符号化方式を用いた符号化方法であれば、ステップ S 1 2 1 2 に進んで復号するが、それ以外の場合には、ステップ S 1 2 0 9 に進む。ステップ S 1 2 0 9 では、ステップ S 1 2 0 5 で読み出した符号化情報に基づいて、ステップ S 1 2 0 6 で読み出した素片データの符号化方法を調べる。ここで μ -law 方式を用いた符号化方法であれば、ステップ S 1 2 1 3 に進んで復号する。それ以外の場合には、ステップ S 1 2 1 0 に進み、ステップ S 1 2 0 5 で読み出した符号化情報に基づいて、ステップ S 1 2 0 6 で読み出した素片データの符号化方法を調べる。スカラ量子化方式を用いた符号化方法であれば、ステップ S 1 2 1 4 に進んで復号する。それ以外の場合には、ステップ S 1 2 1 1 に進み、ステップ S 1 2 0 4 で求めた素片系列に含まれる全ての音声素片に対応する素片データを復号化したか否かを判定する。全ての素片データを復号化した場合には、ステップ S 1 2 1 5 に進み、まだ復号化していない素片データがある場合には、次の素片データを復号するためにステップ S 1 2 0 4 に進む。

【 0 1 5 2 】

ステップ S 1 2 0 7 で符号化されていない場合、或はステップ S 1 2 1 2 ~ S 1 2 1 4 のいずれかで復号されるか、或はステップ S 1 2 1 1 で次の素片データが無くなるとステップ S 1 2 1 5 に進み、ステップ S 1 2 0 3 で決定した韻律に基づいて、復号化した音声素片のそれぞれを変形して接続する（即ち、波形編集する）。次にステップ S 1 2 1 6 に進み、ステップ S 1 2 1 5 で求めた合成音声を出力装置 1 0 3 の有するスピーカから出力する。

【 0 1 5 3 】

以上説明したように実施の形態 5 の音声合成アルゴリズムによれば、所望の音声素片を μ -law 方式、スカラ量子化方式、線形予測符号化方式の何れかに対応する復号方法によって復号することができるので、自然で高品質な合成音声を生成することができる。

【 0 1 5 4 】

なお、実施の形態 5 では、上述の音声合成アルゴリズムを記憶装置 1 0 1 に記憶されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【 0 1 5 5 】

〔実施の形態 6〕

次に、図 1 の音声処理装置を用いて本発明の実施の形態 6 における素片辞書作成アルゴリズム及び音声合成アルゴリズムについて説明する。

【 0 1 5 6 】

前述の実施の形態 5 では、素片辞書 1 1 2 に登録する素片データごとに、符号化方式の異なる複数の符号化方法の中から最適な符号化方法を選択する例について説明した。これに対して、実施の形態 6 では、素片データの種類ごと、符号化方式の異なる複数の符号化方法の中から最適な符号化方法を選択する例について説明する。なお、素片辞書 1 1 2 に登録する音声素片は、音素、半音素、ダイフオン（CV，VC 等），VCV（またはCVC），これらの組み合わせ等からなる。

【 0 1 5 7 】

（素片辞書作成）

図 1 3 は、本発明の実施の形態 6 における素片辞書作成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置 1 0 1 に記憶されている。中央処理装置 1 0 0 は、ユーザの指示に基づいて記憶装置 1 0 1 からこのプログラムを読み出し、以下に示す手順を実行する。

【 0 1 5 8 】

まずステップ S 1 3 0 1 では、外部記憶装置 1 0 2 の素片データ 1 1 1 に記憶されている N 個の素片データ（各素片データは非圧縮状態にある）のそれぞれを指示するインデックス i を、“0”に初期化する。なお、このインデックス i は記憶装置 1 0 1 に設定されている。次にステップ S 1 3 0 2 に進み、インデックス i が指示する i 番目の素片データ Wi を読み出す。こうして読込まれたデータを、

$$W_i = \{x_0, x_1, \dots, x_{T-1}\}$$

とする。ここで、 T は W_i の時間長（単位はサンプル）である。

【0159】

次にステップS1303に進み、ステップS1302で読み出した素片データ W_i の種類を判別する。具体的には、素片データ W_i の種類が、有声摩擦音、破裂音、無声音、鼻音、その他の有声音の何れかであるかを判別する。ここで、素片データ W_i の種類が有声摩擦音、破裂音の何れかである場合はステップS1316に進み、素片データ W_i を非圧縮とする。このように構成することにより、有声摩擦音や破裂音の品質の劣化を防止することができる。このステップS1316では、素片データ W_i の符号化情報を素片辞書112に書き込む。この符号化情報には、素片データ W_i の種類、素片データ W_i が符号化されていないことを示す情報などが含まれている。次にステップS1317に進み、符号化せずに素片データ W_i を素片辞書112に書き込み、ステップS1318に進む。

【0160】

ステップS1305で、素片データ W_i の種類が無声音である場合はステップS1306に進み、素片データ W_i を前述の実施の形態2又は3で説明した符号化方式（即ち、スカラ量子化方式）で符号化する。次にステップS1307に進み、素片データ W_i の符号化情報を素片辞書112に書き込む。この符号化情報には、素片データ W_i の種類、素片データ W_i を符号化した符号化方法を特定する情報、素片データ W_i の復号に必要な情報（量子化コードブック等）などが含まれている。次にステップS1308に進み、ステップS1306で符号化された素片データ W_i を素片辞書112に書き込んでステップS1318に進む。

【0161】

ステップS1309で、素片データ W_i の種類が鼻音である場合はステップS1310に進み、素片データ W_i を前述の実施の形態4で説明した符号化方式（即ち、線形予測符号化方式）で符号化する。次にステップS1311に進み、素片データ W_i の符号化情報を素片辞書112に書き込む。この符号化情報には、素片データ W_i の種類、素片データ W_i を符号化した符号化方法を特定する情報、素片データ W_i の復号に必要な情報（予測係数、量子化コードブック等）などが

含まれている。そして次にステップS1312に進み、ステップS1310で符号化された素片データWiを素片辞書112に書き込んでステップS1318に進む。

【0162】

またステップS1309で、素片データWiの種類がその他の有声音である場合はステップS1313に進み、前述の実施の形態1で説明した符号化方式（即ち、7ビット μ -law方式、又は8ビット μ -law方式）で符号化する。そしてステップS1314に進み、素片データWiの符号化情報を素片辞書112に書き込む。この符号化情報には、素片データWiの種類、素片データWiを符号化した符号化方法を特定する情報、素片データWiの復号に必要な情報などが含まれている。次にステップS1315に進み、ステップS1313で符号化された素片データWiを素片辞書112に書き込んでステップS1318に進む。

【0163】

そしてステップS1318では、N個の素片データの全てに対して前述の処理を行ったか否かを判定する。ここで、 $i = N - 1$ が成立すれば本アルゴリズムを終了する。一方、そうでなければステップS1319に進んで、インデックスiに1を加え、再びステップS1302に進み、更新されたインデックスiが指定する素片データを読み出す。このような処理をN個の素片データの全てに対して繰り返し実行する。

【0164】

以上説明したように実施の形態6の素片辞書作成アルゴリズムによれば、素片辞書112に登録する音声素片の種類ごとに、 μ -law方式で符号化するか、スカラー量子化方式で符号化するか、線形予測符号化方式で符号化するかを選択することができる。このように構成することにより、素片辞書に登録する音声素片の品質を劣化させることなく、素片辞書に必要な記憶容量を非常に効率的に削減することができる。また、従来と同程度の記憶容量をもつ素片辞書に、従来よりも多くの種類の音声素片に登録することもできる。

【0165】

なお実施の形態6では、上述の素片辞書作成アルゴリズムを記憶装置101に

記憶されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【0166】

（音声合成）

図14は、本発明の実施の形態6における音声合成アルゴリズムを説明するフローチャートである。このアルゴリズムを実現するプログラムは記憶装置101に記憶されている。中央処理装置100は、ユーザの指示に基づいて記憶装置101からこのプログラムを読み出し、以下に示す手順を実行する。

【0167】

ステップS1401～S1403のそれぞれは、図12のS1201～S1203のそれぞれと同様の機能及び処理をするものとし、その説明を省略する。

【0168】

ステップS1404では、ステップS1402で求めた音韻系列とステップS1403で決定した韻律とに基づいて最適な素片系列を求める。そして、この素片系列に含まれる音声素片の一つを選択し、選択された音声素片に対応する素片データ、符号化情報を検索する。そして、素片辞書112がハードディスク等の記憶媒体に格納されているならば、素片データ、符号化情報の記憶領域へ順次シークする。また、素片辞書112がRAM等の記憶媒体に格納されているならば、素片データ、符号化情報の記憶領域へポインタ（アドレスレジスタ）を順次移動する。

【0169】

ステップS1405では、ステップS1404で検索した符号化情報を素片辞書112から読み出す。次にステップS1406に進み、ステップS1404で検索した素片データを素片辞書112から読み出す。次にステップS1407に進み、ステップS1405で読み出した符号化情報に基づいて、ステップS1404で検索された素片データの種別を判別する。具体的には、素片データの種別が、有声摩擦音、破裂音、無声音、鼻音、その他の有声音の何れかであるかを判別する。

【0170】

ここで素片データの種類の種類が有声摩擦音、破裂音の何れかである場合はステップ S 1 4 1 6 に進み、ステップ S 1 4 0 4 で検索された素片データを読み出し、ステップ S 1 4 1 7 に進む。この場合、この素片データは符号化されていない。

【0171】

ステップ S 1 4 0 7 で、素片データの種類の種類が有声摩擦音、破裂音の何れかでない場合はステップ S 1 4 0 8 に進み、素片データの種類の種類が無声音であるかどうかを調べ、無声音であるときはステップ S 1 4 1 4 に進む。ステップ S 1 4 1 4 では、ステップ S 1 4 0 4 で検索された素片データを読み出してステップ S 1 4 1 5 に進む。この素片データはスカラ量子化方式で符号化されている。ステップ S 1 4 1 5 では、ステップ S 1 4 0 5 で読み出した符号化情報に基づいてこの素片データを復号する。

【0172】

ステップ S 1 4 0 8 で素片データの種類の種類が無声音でない場合はステップ S 1 4 0 9 に進み、素片データの種類の種類が鼻音であるかどうかを判定し、鼻音であればステップ S 1 4 1 2 に進み、ステップ S 1 4 0 4 で検索された素片データを読み出してステップ S 1 4 1 3 に進む。この素片データは線形予測符号化方式で符号化されている。ステップ S 1 4 1 3 では、ステップ S 1 4 0 5 で読み出した符号化情報に基づいてこの素片データを復号する。

【0173】

一方、ステップ S 1 4 0 9 で、素片データの種類の種類がその他の有声音である場合はステップ S 1 4 1 0 に進み、ステップ S 1 4 0 4 で検索された素片データを読み出してステップ S 1 4 1 1 に進む。この素片データは μ -law 方式で符号化されている。ステップ S 1 4 1 1 では、ステップ S 1 4 0 5 で読み出した符号化情報に基づいてこの素片データを復号する。

【0174】

こうしてステップ S 1 4 1 7 では、ステップ S 1 4 0 4 で求めた素片系列に含まれる全ての音声素片に対応する素片データを復号化したか否かを判定する。全ての素片データを復号化した場合はステップ S 1 4 1 8 に進み、まだ復号化して

いない素片データがある場合には、次の素片データを復号するためにステップ S 1 4 0 4 に進む。ステップ S 1 4 1 8 では、ステップ S 1 4 0 3 で決定した韻律に基づいて、復号化した音声素片のそれぞれを変形して接続する（即ち、波形編集する）。次にステップ S 1 4 1 9 に進み、ステップ S 1 4 1 8 で求めた合成音声を出力装置 1 0 3 の有するスピーカから出力する。

【 0 1 7 5 】

以上説明したように実施の形態 6 の音声合成アルゴリズムによれば、所望の音声素片を μ -law 方式、スカラ量子化方式、線形予測符号化方式の何れかに対応する復号方法によって復号することができるので、自然で高品質な合成音声を生成することができる。

【 0 1 7 6 】

なお、実施の形態 6 では、上述の音声合成アルゴリズムを記憶装置 1 0 1 に記憶されているプログラムに基づいて実現する場合について説明したが、この素片辞書作成アルゴリズムの一部若しくは全てをハードウェアにより構成することも可能である。

【 0 1 7 7 】

〔その他の実施の形態〕

前述の実施の形態 2, 3, 4 および 5 では、量子化の手法としてスカラ量子化を用いたが、これに限るものではない。連続する複数のサンプルを 1 つのベクトルと見なし、ベクトル量子化を適用することも可能である。

【 0 1 7 8 】

また、破裂音のような非定常性の音声素片については、音声素片を破裂前と破裂後とに分割し、それぞれを最適な符号化方法で符号化することも可能である。これにより、非定常性の音声素片の符号化効率をより一層向上させることが可能となる。

【 0 1 7 9 】

また、実施の形態 4 では、線形予測モデルに基づいて説明を行ったが、これに限るものではない。その他の声道フィルタモデルを適用することも可能である。例えば、線形予測係数の代わりに LMA (Log Magnitude Approximation) フィ

ルタ係数を用い、予測残差の代わりにLMAフィルタの残差を用いてモデルパラメータを演算することも可能である。このように構成することによって、実施の形態4をケプストラム領域に適用することができる。

【0180】

なお、前述の各実施の形態は、複数の機器（例えばホストコンピュータ、インターフェース機器、リーダー、プリンタなど）から構成されるシステムに適用しても、一つの機器からなる装置（例えば、複写機、ファクシミリ装置など）に適用することも可能である。

【0181】

また、前述の各実施の形態は、中央処理装置100の読み出したプログラムコードの指示に基づき、中央処理装置100上で稼働しているオペレーティングシステム（OS）などが実際の処理の一部または全てを行うように構成することも可能である。

【0182】

さらに、前述の各実施の形態は、記憶装置101から読み出されたプログラムコードを中央処理装置100に接続された機能拡張ユニットに備わるメモリに書き込んだ後、その機能拡張ユニットに備わるCPUなどが、そのプログラムコードの指示に基づき、実際の処理の一部または全部を行うように構成することも可能である。

【0183】

以上説明したように前述の各実施の形態によれば、素片データごとに符号化方法を選択することができるので、素片辞書に登録する音声素片の品質を劣化させることなく、素片辞書に必要な記憶容量を非常に効率的に削減することができる。また、このようにして作成された素片辞書を用いて合成音声を生成することができるので、自然で高品質な合成音声を生成することができる。

【0184】

【発明の効果】

以上説明したように本発明によれば、素片辞書に登録する音声素片の品質を劣化させることなく、素片辞書に必要な記憶容量を非常に効率的に削減することが

できる。

【 0 1 8 5 】

また本発明によれば、自然で高品質な合成音声を生成することができる。

【図面の簡単な説明】

【図 1】

本発明の各実施の形態における音声合成装置のハードウェア構成を示すブロック図である。

【図 2】

本発明の実施の形態 1 における素片辞書作成アルゴリズムを説明するフローチャートである。

【図 3】

本発明の実施の形態 1 における音声合成アルゴリズムを説明するフローチャートである。

【図 4】

本発明の実施の形態 2 における素片辞書作成アルゴリズムを説明するフローチャートである。

【図 5】

本発明の実施の形態 2 における音声合成アルゴリズムを説明するフローチャートである。

【図 6】

本発明の実施の形態 3 における素片辞書作成アルゴリズムを説明するフローチャートである。

【図 7】

本発明の実施の形態 3 における素片辞書作成アルゴリズムを説明するフローチャートである。

【図 8】

本発明の実施の形態 3 における音声合成アルゴリズムを説明するフローチャートである。

【図 9】

本発明の実施の形態 4 における素片辞書作成アルゴリズムを説明するフローチャートである。

【図 1 0】

本発明の実施の形態 4 における音声合成アルゴリズムを説明するフローチャートである。

【図 1 1】

本発明の実施の形態 5 における素片辞書作成アルゴリズムを説明するフローチャートである。

【図 1 2】

本発明の実施の形態 5 における音声合成アルゴリズムを説明するフローチャートである。

【図 1 3】

本発明の実施の形態 6 における素片辞書作成アルゴリズムを説明するフローチャートである。

【図 1 4】

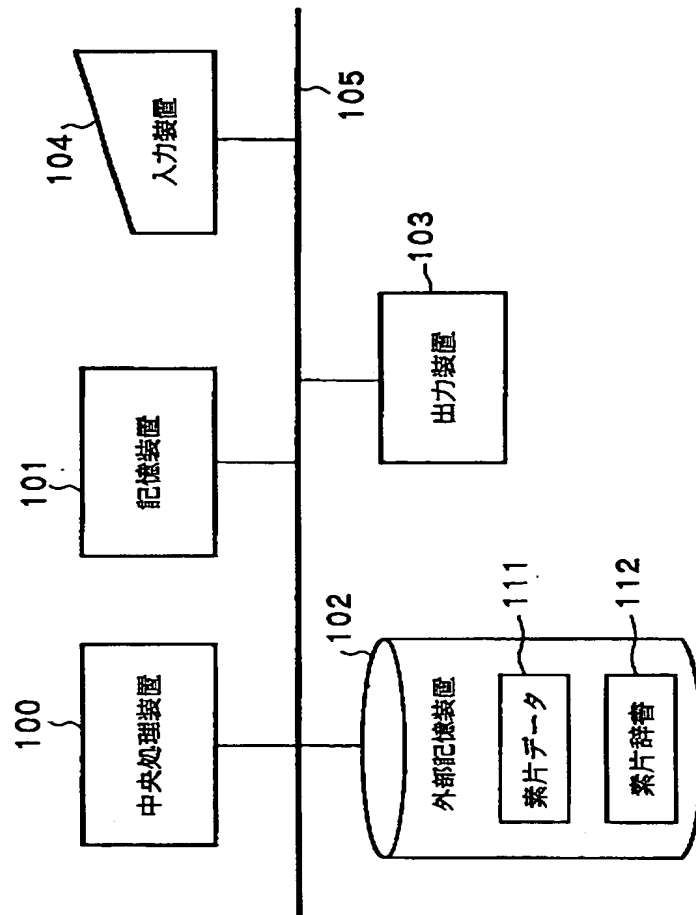
本発明の実施の形態 6 における音声合成アルゴリズムを説明するフローチャートである。

【図 1 5】

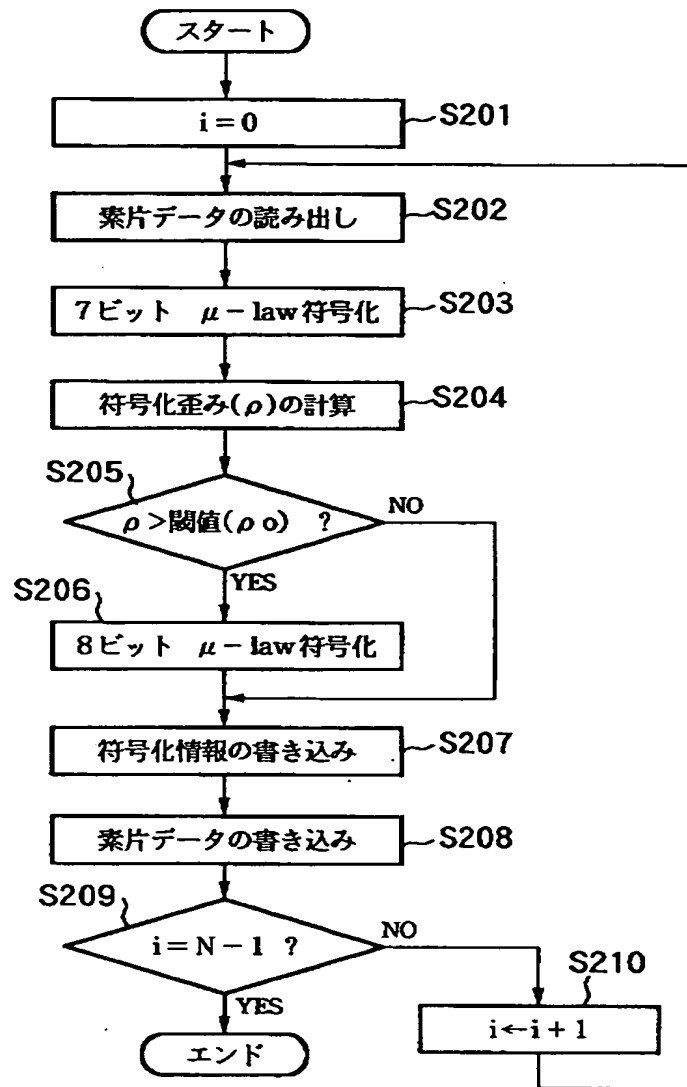
一般的な音声合成処理を示すフローチャートである。

【書類名】 図面

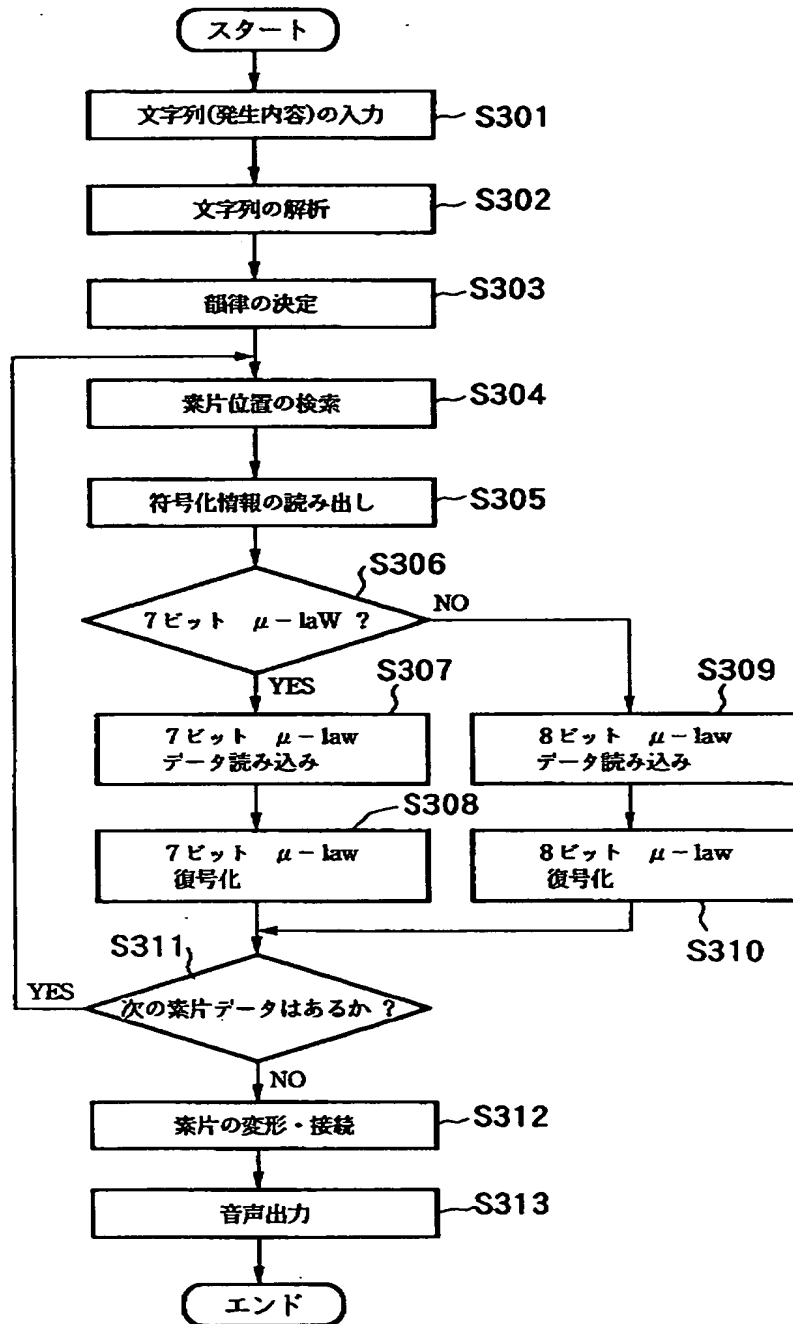
【図 1】



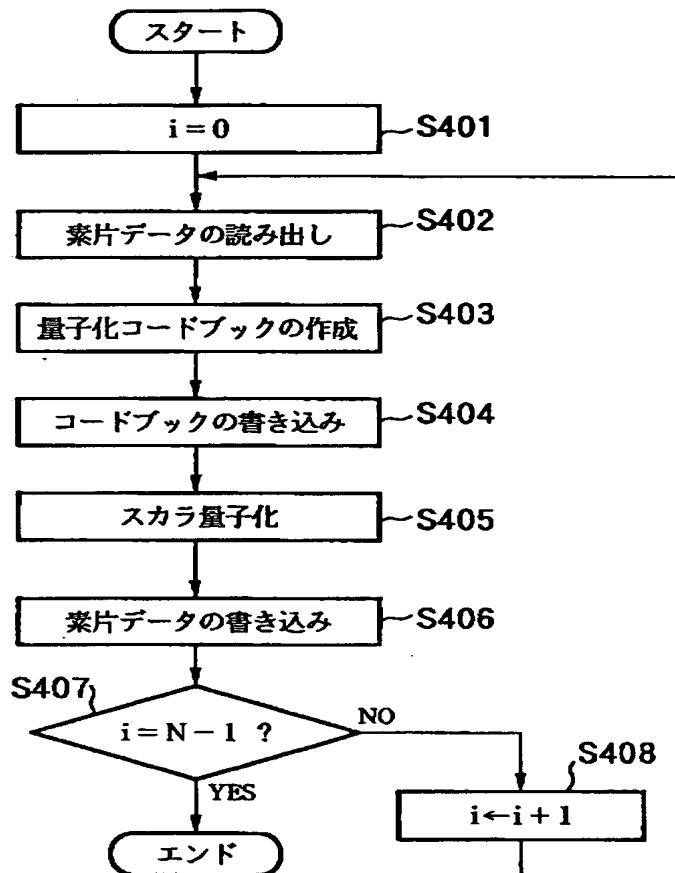
【図 2】



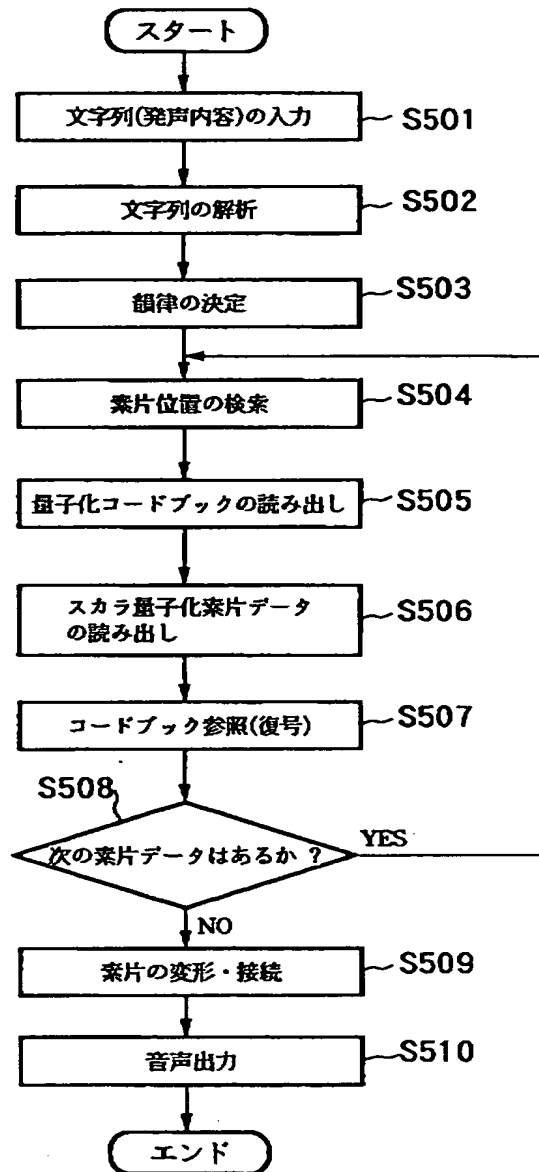
【図 3】



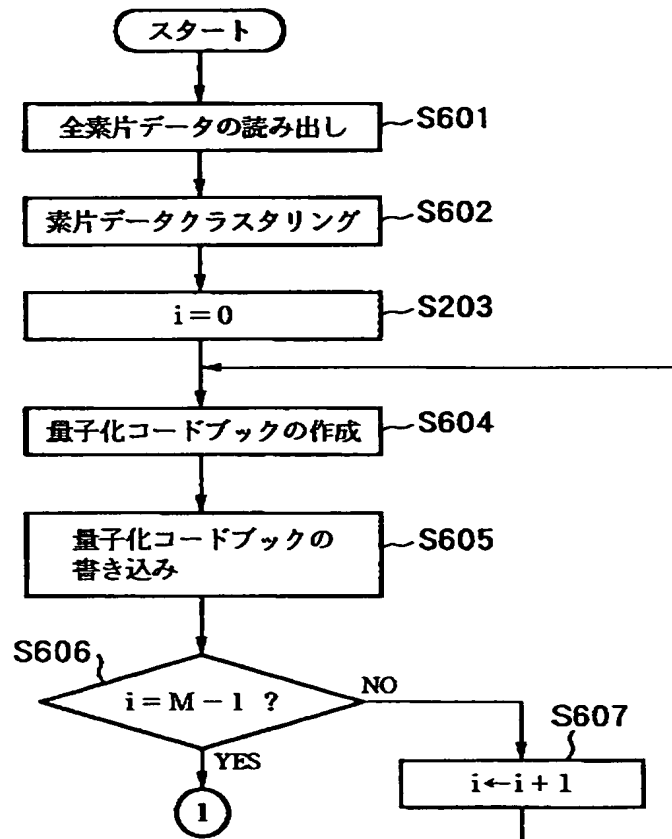
【図 4】



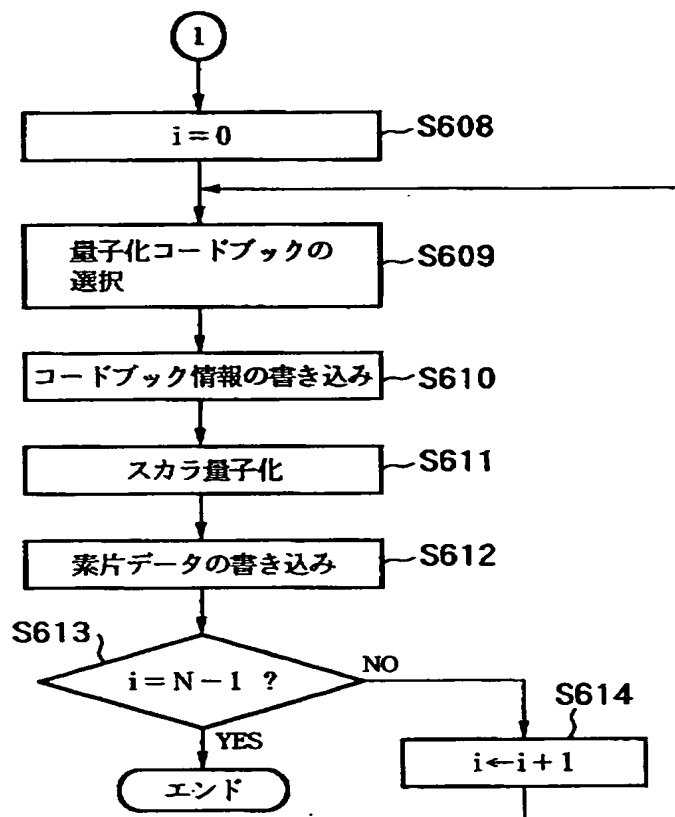
【図 5】



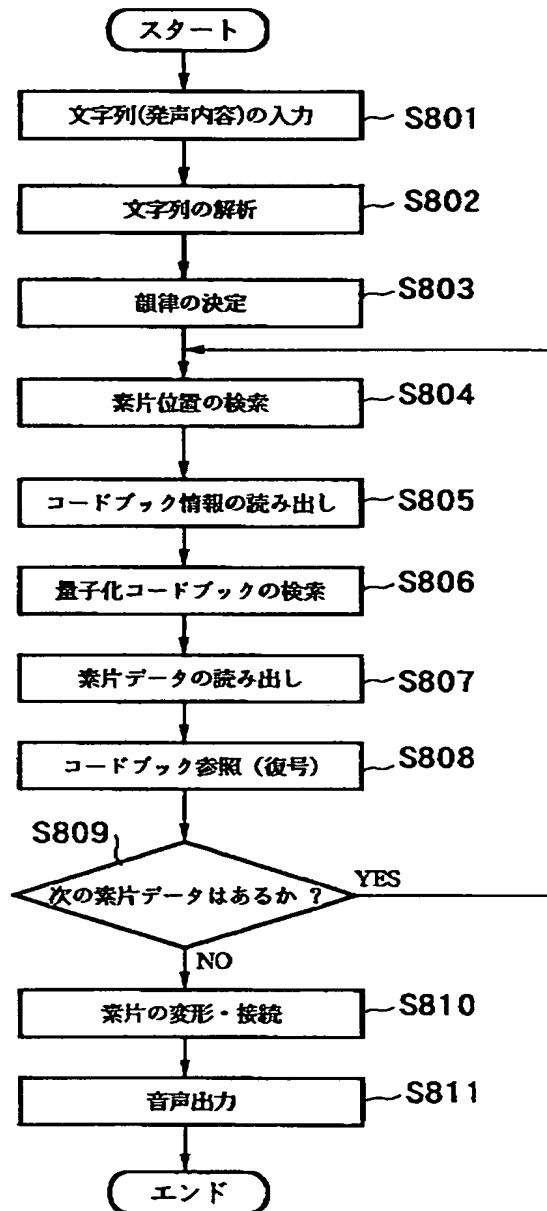
【図 6】



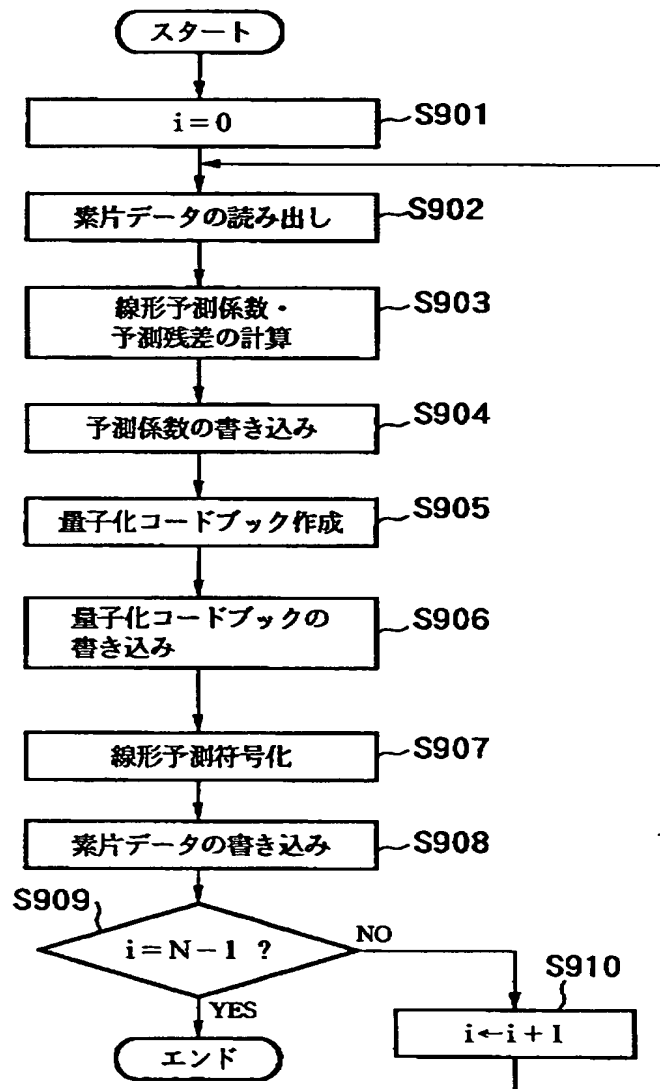
【図 7】



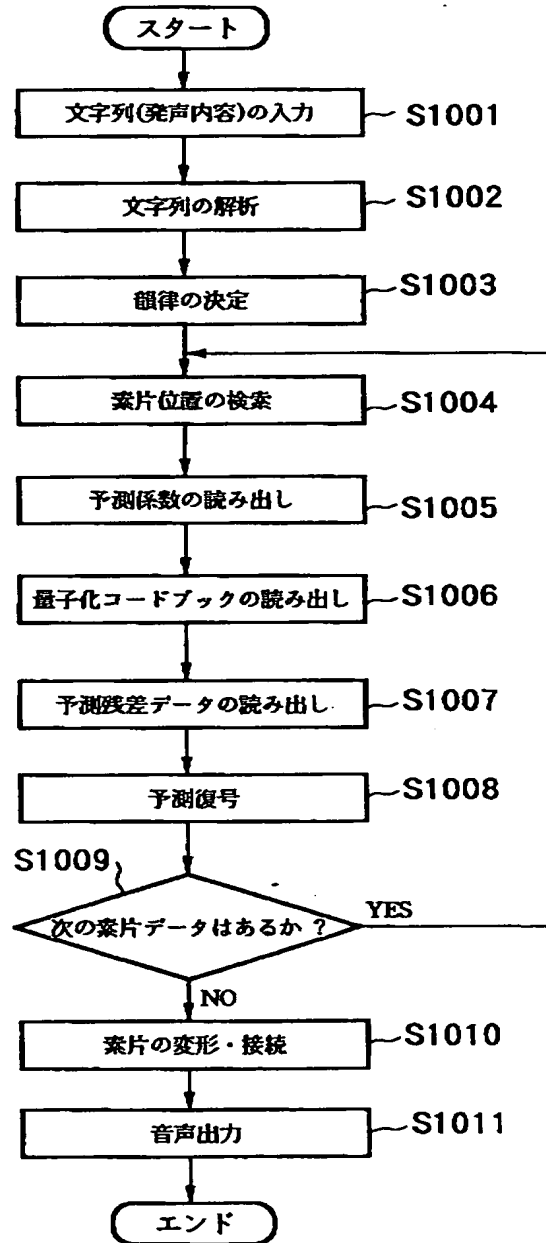
【図 8】



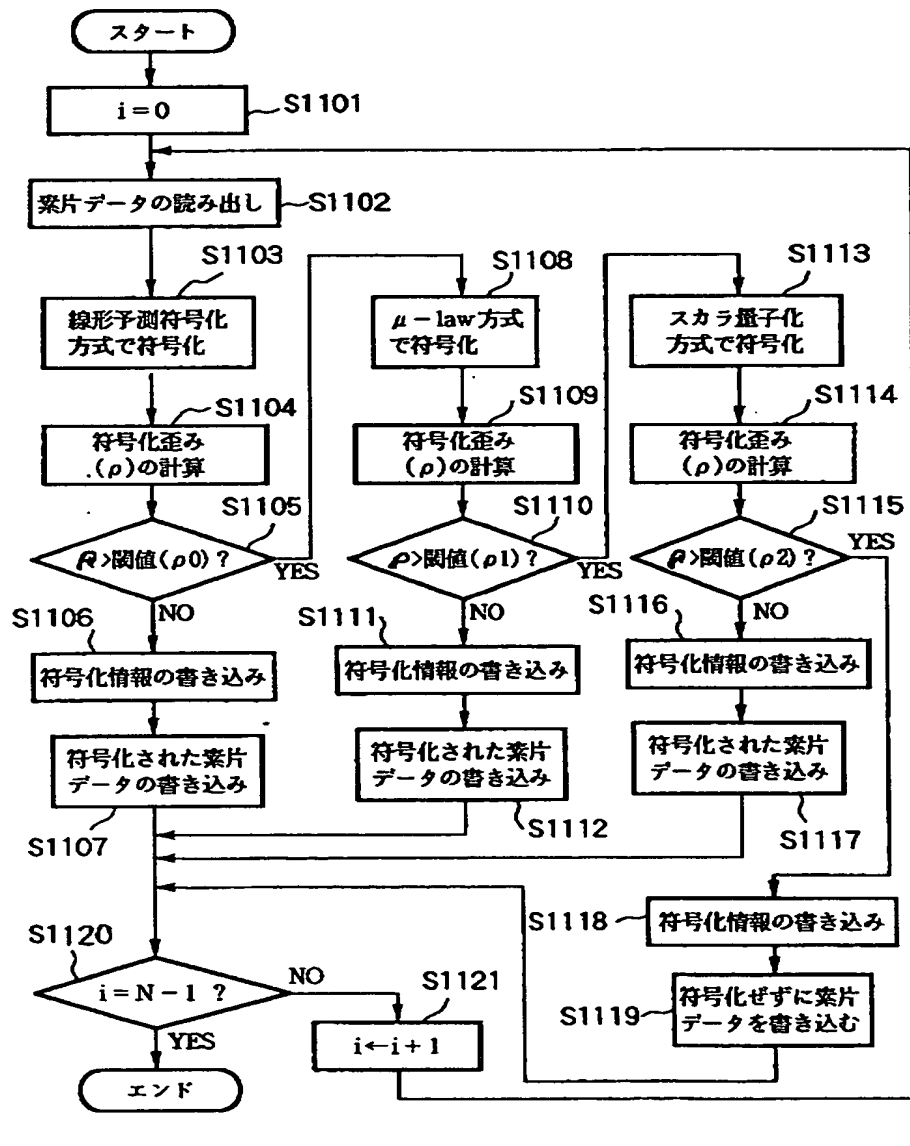
【図 9】



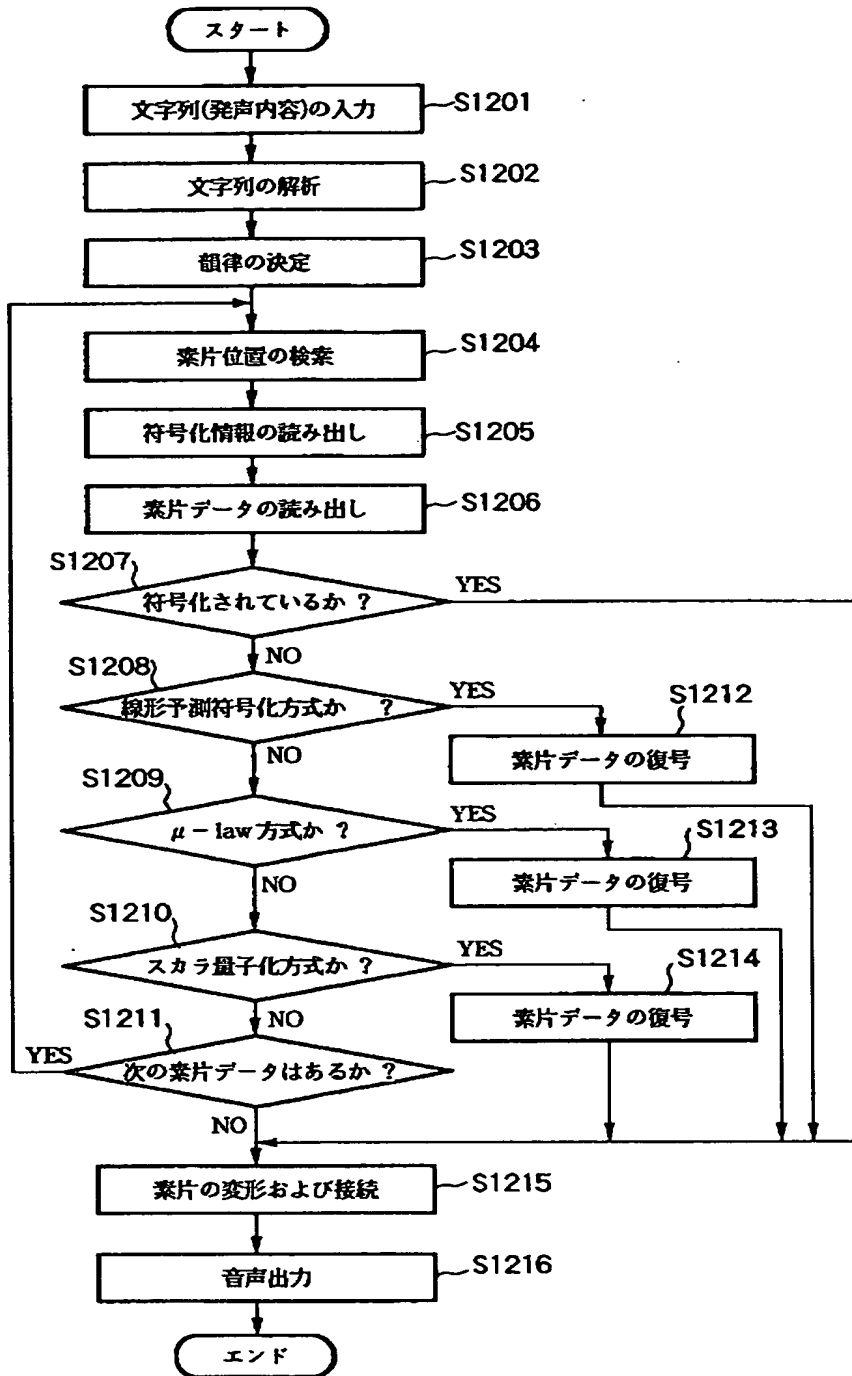
【図 1 0】



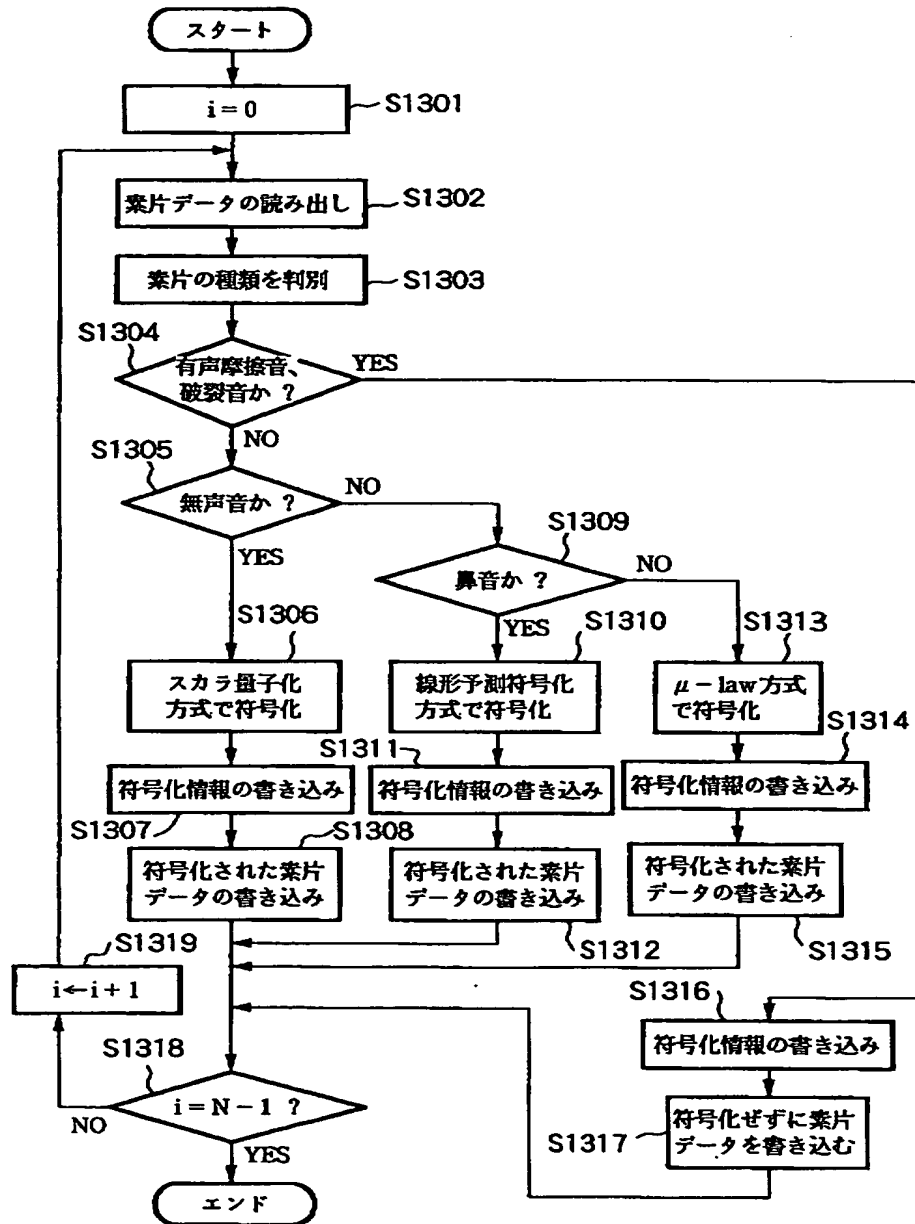
【図11】



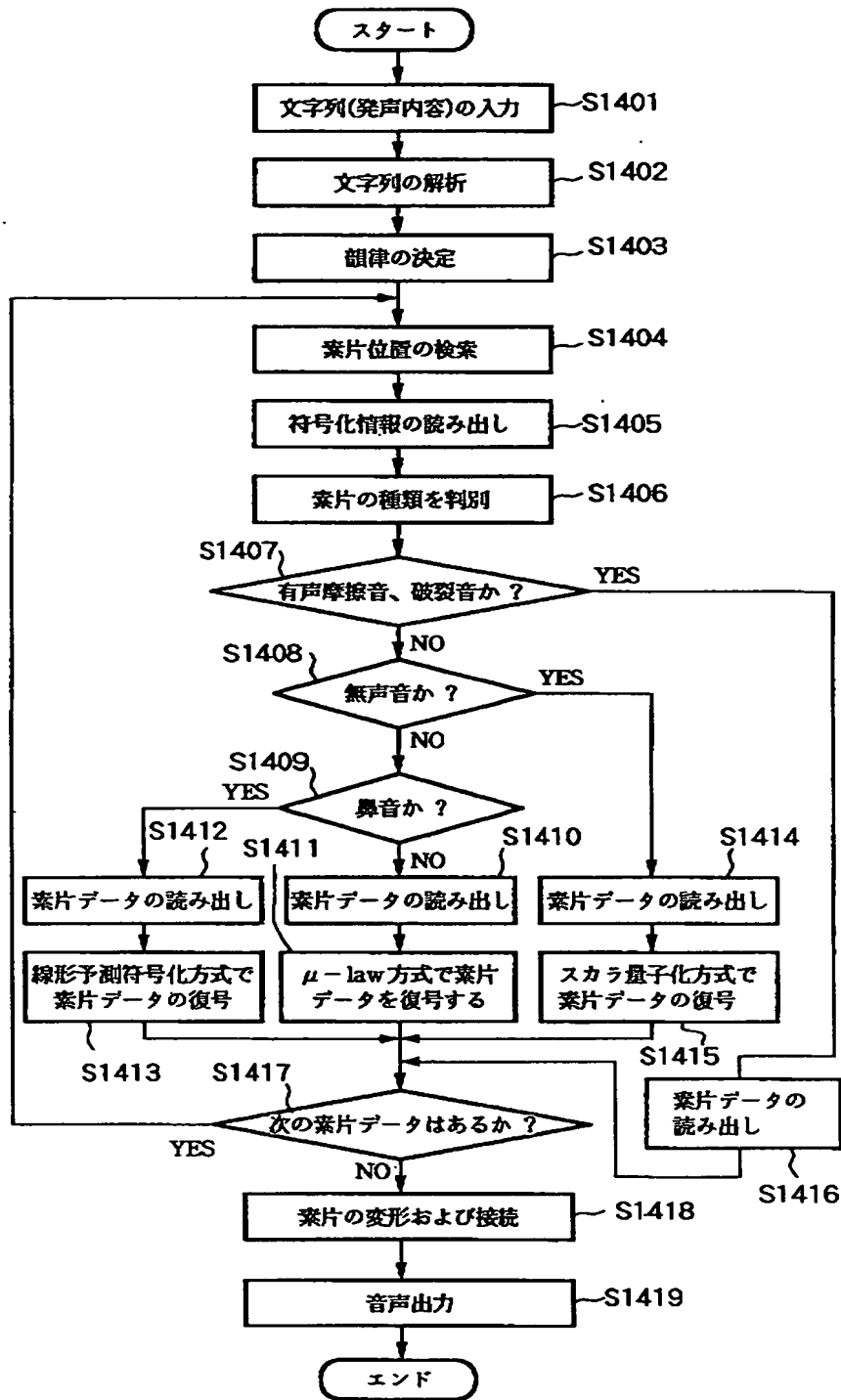
【図 1 2】



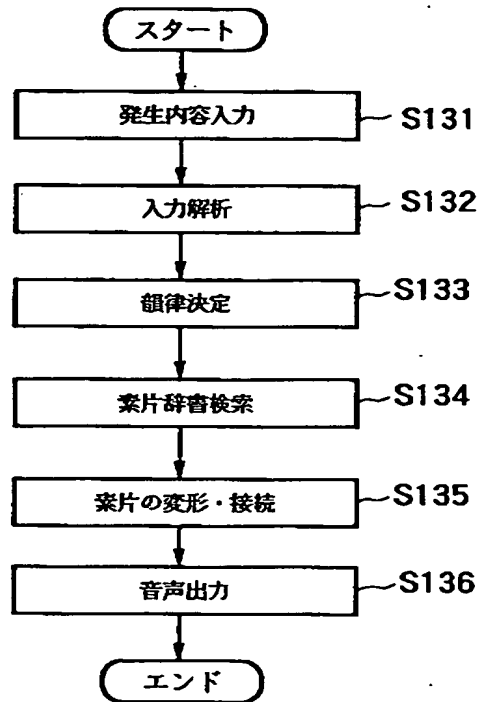
【図 13】



【図 1 4】



【図 1 5】



【書類名】 要約書

【要約】

【課題】 素片辞書に登録する音声素片の品質を劣化させることなく、素片辞書に必要な記憶容量を非常に効率的に削減する。

【解決手段】 素片データ 1 1 1 から N 個の素片データを読み出し、それぞれ最適な符号化方法に従って符号化する。こうして符号化された素片データを、符号化する際に使用した符号化方法を特定する情報とともに素片辞書 1 1 2 に登録する。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願 2 0 0 0 - 2 2 1 1 2 8
受付番号	5 0 0 0 0 9 2 4 6 0 5
書類名	特許願
担当官	第八担当上席 0 0 9 7
作成日	平成 1 2 年 7 月 2 6 日

<認定情報・付加情報>

【特許出願人】

【識別番号】	000001007
【住所又は居所】	東京都大田区下丸子 3 丁目 3 0 番 2 号
【氏名又は名称】	キヤノン株式会社

【代理人】

申請人

【識別番号】	100076428
【住所又は居所】	東京都千代田区紀尾井町 3 番 6 号 秀和紀尾井町 パークビル 7 F 大塚国際特許事務所
【氏名又は名称】	大塚 康德

【選任した代理人】

【識別番号】	100101306
【住所又は居所】	東京都千代田区紀尾井町 3 番 6 号 秀和紀尾井町 パークビル 7 F 大塚国際特許事務所
【氏名又は名称】	丸山 幸雄

【選任した代理人】

【識別番号】	100115071
【住所又は居所】	東京都千代田区紀尾井町 3 番 6 号 秀和紀尾井町 パークビル 7 F 大塚国際特許事務所
【氏名又は名称】	大塚 康弘

出 願 人 履 歴 情 報

識別番号 [000001007]

1. 変更年月日 1990年 8月30日

[変更理由] 新規登録

住 所 東京都大田区下丸子3丁目30番2号
氏 名 キヤノン株式会社